

## COMPOSITIONS AND METHODS FOR THE THERAPY AND DIAGNOSIS OF LUNG CANCER

### CROSS REFERENCE TO RELATED APPLICATIONS

5 This application is related to U.S. Provisional Applications Nos. 60/196,780, filed April 11, 2000; 60/213,361, filed June 21, 2000; 60/229,763, filed September 1, 2000; 60/230,629, filed September 5, 2000; 60/232,565, filed Septemer 14, 2000; 60/257,037, filed December 19, 2000; and 60/260,796, filed January 8, 2001, all of which are incorporated in their entirety herein.

### 10 TECHNICAL FIELD OF THE INVENTION

The present invention relates generally to therapy and diagnosis of cancer, such as lung cancer. The invention is more specifically related to polypeptides, comprising at least a portion of a lung tumor protein, and to polynucleotides encoding such polypeptides. Such polypeptides and polynucleotides are useful in pharmaceutical compositions, *e.g.*, vaccines, and other compositions for the diagnosis and treatment of lung cancer.

### BACKGROUND OF THE INVENTION

Lung cancer is the primary cause of cancer death among both men and women in the U.S., with an estimated 172,000 new cases being reported in 1994. The five-year survival rate among all lung cancer patients, regardless of the stage of disease at diagnosis, is only 13%. This contrasts with a five-year survival rate of 46% among cases detected while the disease is still localized. However, only 16% of lung cancers are discovered before the disease has spread.

Early detection is difficult since clinical symptoms are often not seen until 25 the disease has reached an advanced stage. Currently, diagnosis is aided by the use of chest x-rays, analysis of the type of cells contained in sputum and fiberoptic examination of the

bronchial passages. Treatment regimens are determined by the type and stage of the cancer, and include surgery, radiation therapy and/or chemotherapy. In spite of considerable research into therapies for the disease, lung cancer remains difficult to treat.

Accordingly, there remains a need in the art for improved vaccines,  
5 treatment methods and diagnostic techniques for lung cancer.

## SUMMARY OF THE INVENTION

In one aspect, the present invention provides polynucleotide compositions comprising a sequence selected from the group consisting of:

- (a) sequences provided in SEQ ID NO:1-232, 243-396, 398-412, 414-424 and 437-440;
- 10 (b) complements of the sequences provided in SEQ ID NO:1-232, 243-396, 398-412, 414-424 and 437-440;
- (c) sequences consisting of at least 20 contiguous residues of a sequence provided in SEQ ID NO:1-232, 243-396, 398-412, 414-424 and 437-440;
- 15 (d) sequences that hybridize to a sequence provided in SEQ ID NO:1-232, 243-396, 398-412, 414-424 and 437-440, under moderately stringent conditions;
- (e) sequences having at least 75% identity to a sequence of SEQ ID NO:1-232, 243-396, 398-412, 414-424 and 437-440;
- 20 (f) sequences having at least 90% identity to a sequence of SEQ ID NO:1-232, 243-396, 398-412, 414-424 and 437-440; and
- (g) degenerate variants of a sequence provided in SEQ ID NO:1-232, 243-396, 398-412, 414-424 and 437-440.

In one preferred embodiment, the polynucleotide compositions of the invention are expressed in at least about 20%, more preferably in at least about 30%, and most preferably in at least about 50% of lung tumor samples tested, at a level that is at least about 2-fold, preferably at least about 5-fold, and most preferably at least about 10-fold higher than that for normal tissues.

The present invention, in another aspect, provides polypeptide compositions comprising an amino acid sequence that is encoded by a polynucleotide sequence described above. In certain specific embodiments, the present invention provides polypeptide compositions comprising an amino acid sequence selected from the group consisting of 5 sequences recited in SEQ ID NO:229-232, 237-242, 397, 413 and 425-436.

In certain preferred embodiments, the polypeptides and/or polynucleotides of the present invention are immunogenic, *i.e.*, they are capable of eliciting an immune response, particularly a humoral and/or cellular immune response, as further described herein.

10 The present invention further provides fragments, variants and/or derivatives of the disclosed polypeptide and/or polynucleotide sequences, wherein the fragments, variants and/or derivatives preferably have a level of immunogenic activity of at least about 50%, preferably at least about 70% and more preferably at least about 90% of the level of 15 immunogenic activity of a polypeptide sequence set forth in SEQ ID NOs: 229-232, 237-242, 397, 413 and 425-436, or a polypeptide sequence encoded by a polynucleotide sequence set forth in SEQ ID NOs: 1-232, 243-396, 398-412, 414-424 and 437-440.

The present invention further provides polynucleotides that encode a polypeptide described above, expression vectors comprising such polynucleotides and host cells transformed or transfected with such expression vectors.

20 Within other aspects, the present invention provides pharmaceutical compositions comprising a polypeptide or polynucleotide as described above and a physiologically acceptable carrier.

Within a related aspect of the present invention, the pharmaceutical compositions, *e.g.*, vaccine compositions, are provided for prophylactic or therapeutic 25 applications. Such compositions generally comprise an immunogenic polypeptide or polynucleotide of the invention and an immunostimulant, such as an adjuvant.

The present invention further provides pharmaceutical compositions that comprise: (a) an antibody or antigen-binding fragment thereof that specifically binds to a

polypeptide of the present invention, or a fragment thereof; and (b) a physiologically acceptable carrier.

Within further aspects, the present invention provides pharmaceutical compositions comprising: (a) an antigen presenting cell that expresses a polypeptide as described above and (b) a pharmaceutically acceptable carrier or excipient. Illustrative antigen presenting cells include dendritic cells, macrophages, monocytes, fibroblasts and B cells.

Within related aspects, pharmaceutical compositions are provided that comprise: (a) an antigen presenting cell that expresses a polypeptide as described above and (b) an immunostimulant.

The present invention further provides, in other aspects, fusion proteins that comprise at least one polypeptide as described above, as well as polynucleotides encoding such fusion proteins, typically in the form of pharmaceutical compositions, e.g., vaccine compositions, comprising a physiologically acceptable carrier and/or an immunostimulant.

The fusions proteins may comprise multiple immunogenic polypeptides or portions/variants thereof, as described herein, and may further comprise one or more polypeptide segments for facilitating the expression, purification and/or immunogenicity of the polypeptide(s).

Within further aspects, the present invention provides methods for stimulating an immune response in a patient, preferably a T cell response in a human patient, comprising administering a pharmaceutical composition described herein. The patient may be afflicted with lung cancer, in which case the methods provide treatment for the disease, or patient considered at risk for such a disease may be treated prophylactically.

Within further aspects, the present invention provides methods for inhibiting the development of a cancer in a patient, comprising administering to a patient a pharmaceutical composition as recited above. The patient may be afflicted with lung cancer, in which case the methods provide treatment for the disease, or patient considered at risk for such a disease may be treated prophylactically.

The present invention further provides, within other aspects, methods for removing tumor cells from a biological sample, comprising contacting a biological sample

with T cells that specifically react with a polypeptide of the present invention, wherein the step of contacting is performed under conditions and for a time sufficient to permit the removal of cells expressing the protein from the sample.

Within related aspects, methods are provided for inhibiting the development 5 of a cancer in a patient, comprising administering to a patient a biological sample treated as described above.

Methods are further provided, within other aspects, for stimulating and/or expanding T cells specific for a polypeptide of the present invention, comprising contacting T cells with one or more of: (i) a polypeptide as described above; (ii) a polynucleotide 10 encoding such a polypeptide; and/or (iii) an antigen presenting cell that expresses such a polypeptide; under conditions and for a time sufficient to permit the stimulation and/or expansion of T cells. Isolated T cell populations comprising T cells prepared as described above are also provided.

Within further aspects, the present invention provides methods for inhibiting 15 the development of a cancer in a patient, comprising administering to a patient an effective amount of a T cell population as described above.

The present invention further provides methods for inhibiting the development of a cancer in a patient, comprising the steps of: (a) incubating CD4<sup>+</sup> and/or CD8<sup>+</sup> T cells isolated from a patient with one or more of: (i) a polypeptide comprising at 20 least an immunogenic portion of polypeptide disclosed herein; (ii) a polynucleotide encoding such a polypeptide; and (iii) an antigen-presenting cell that expressed such a polypeptide; and (b) administering to the patient an effective amount of the proliferated T cells, and thereby inhibiting the development of a cancer in the patient. Proliferated cells may, but need not, be cloned prior to administration to the patient.

25 Within further aspects, the present invention provides methods for determining the presence or absence of a cancer, preferably a lung cancer, in a patient comprising: (a) contacting a biological sample obtained from a patient with a binding agent that binds to a polypeptide as recited above; (b) detecting in the sample an amount of polypeptide that binds to the binding agent; and (c) comparing the amount of polypeptide

with a predetermined cut-off value, and therefrom determining the presence or absence of a cancer in the patient. Within preferred embodiments, the binding agent is an antibody, more preferably a monoclonal antibody.

The present invention also provides, within other aspects, methods for monitoring the progression of a cancer in a patient. Such methods comprise the steps of: (a) contacting a biological sample obtained from a patient at a first point in time with a binding agent that binds to a polypeptide as recited above; (b) detecting in the sample an amount of polypeptide that binds to the binding agent; (c) repeating steps (a) and (b) using a biological sample obtained from the patient at a subsequent point in time; and (d) comparing the amount of polypeptide detected in step (c) with the amount detected in step (b) and therefrom monitoring the progression of the cancer in the patient.

The present invention further provides, within other aspects, methods for determining the presence or absence of a cancer in a patient, comprising the steps of: (a) contacting a biological sample obtained from a patient with an oligonucleotide that hybridizes to a polynucleotide that encodes a polypeptide of the present invention; (b) detecting in the sample a level of a polynucleotide, preferably mRNA, that hybridizes to the oligonucleotide; and (c) comparing the level of polynucleotide that hybridizes to the oligonucleotide with a predetermined cut-off value, and therefrom determining the presence or absence of a cancer in the patient. Within certain embodiments, the amount of mRNA is detected via polymerase chain reaction using, for example, at least one oligonucleotide primer that hybridizes to a polynucleotide encoding a polypeptide as recited above, or a complement of such a polynucleotide. Within other embodiments, the amount of mRNA is detected using a hybridization technique, employing an oligonucleotide probe that hybridizes to a polynucleotide that encodes a polypeptide as recited above, or a complement of such a polynucleotide.

In related aspects, methods are provided for monitoring the progression of a cancer in a patient, comprising the steps of: (a) contacting a biological sample obtained from a patient with an oligonucleotide that hybridizes to a polynucleotide that encodes a polypeptide of the present invention; (b) detecting in the sample an amount of a

polynucleotide that hybridizes to the oligonucleotide; (c) repeating steps (a) and (b) using a biological sample obtained from the patient at a subsequent point in time; and (d) comparing the amount of polynucleotide detected in step (c) with the amount detected in step (b) and therefrom monitoring the progression of the cancer in the patient.

5 Within further aspects, the present invention provides antibodies, such as monoclonal antibodies, that bind to a polypeptide as described above, as well as diagnostic kits comprising such antibodies. Diagnostic kits comprising one or more oligonucleotide probes or primers as described above are also provided.

10 These and other aspects of the present invention will become apparent upon reference to the following detailed description and attached drawings. All references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

#### BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE IDENTIFIERS

15 Figure 1 is a bar graph showing expression of clone SCC2-51 in normal tissues and tumor tissues.

SEQ ID NO:1 is the determined cDNA sequence for LSC-1.

SEQ ID NO:2 is the determined cDNA sequence for LSC-2.

SEQ ID NO:3 is the determined cDNA sequence for LSC-3.

SEQ ID NO:4 is the determined cDNA sequence for LSC-5.

20 SEQ ID NO:5 is the determined cDNA sequence for LSC-6.

SEQ ID NO:6 is the determined cDNA sequence for LSC-7.

SEQ ID NO:7 is the determined cDNA sequence for LSC-9.

SEQ ID NO:8 is the determined cDNA sequence for LSC-10.

SEQ ID NO:9 is the determined cDNA sequence for LSC-11.

25 SEQ ID NO:10 is the determined cDNA sequence for LSC-13.

SEQ ID NO:11 is the determined cDNA sequence for LSC-15.

SEQ ID NO:12 is the determined cDNA sequence for LSC-20.

SEQ ID NO:13 is the determined cDNA sequence for LSC-23.

SEQ ID NO:14 is the determined cDNA sequence for LSC-24.  
SEQ ID NO:15 is the determined cDNA sequence for LSC-25.  
SEQ ID NO:16 is the determined cDNA sequence for LSC-26.  
SEQ ID NO:17 is the determined cDNA sequence for LSC-27.  
5       SEQ ID NO:18 is the determined cDNA sequence for LSC-28.  
SEQ ID NO:19 is the determined cDNA sequence for LSC-29.  
SEQ ID NO:20 is the determined cDNA sequence for LSC-30.  
SEQ ID NO:21 is the determined cDNA sequence for LSC-31.  
SEQ ID NO:22 is the determined cDNA sequence for LSC-33.  
10      SEQ ID NO:23 is the determined cDNA sequence for LSC-34.  
SEQ ID NO:24 is the determined cDNA sequence for LSC-35.  
SEQ ID NO:25 is the determined cDNA sequence for LSC-37.  
SEQ ID NO:26 is the determined cDNA sequence for LSC-39.  
SEQ ID NO:27 is the determined cDNA sequence for LSC-43.  
15      SEQ ID NO:28 is the determined cDNA sequence for LSC-46.  
SEQ ID NO:29 is the determined cDNA sequence for LSC-49.  
SEQ ID NO:30 is the determined cDNA sequence for LSC-51.  
SEQ ID NO:31 is the determined cDNA sequence for LSC-53.  
SEQ ID NO:32 is the determined cDNA sequence for LSC-55.  
20      SEQ ID NO:33 is the determined cDNA sequence for LSC-60.  
SEQ ID NO:34 is the determined cDNA sequence for LSC-62.  
SEQ ID NO:35 is the determined cDNA sequence for LSC-64.  
SEQ ID NO:36 is the determined cDNA sequence for LSC-65.  
SEQ ID NO:37 is the determined cDNA sequence for LSC-71.  
25      SEQ ID NO:38 is the determined cDNA sequence for LSC-72.  
SEQ ID NO:39 is the determined cDNA sequence for LSC-74.  
SEQ ID NO:40 is the determined cDNA sequence for LSC-76.  
SEQ ID NO:41 is the determined cDNA sequence for LSC-77.  
SEQ ID NO:42 is the determined cDNA sequence for LSC-78.

SEQ ID NO:43 is the determined cDNA sequence for LSC-81.  
SEQ ID NO:44 is the determined cDNA sequence for LSC-93.  
SEQ ID NO:45 is the determined cDNA sequence for LSC-101.  
SEQ ID NO:46 is the determined cDNA sequence for LSC-102.  
5 SEQ ID NO:47 is the determined cDNA sequence for LSC-103.  
SEQ ID NO:48 is the determined cDNA sequence for LSC-105.  
SEQ ID NO:49 is the determined cDNA sequence for LSC-110.  
SEQ ID NO:50 is the determined cDNA sequence for LSC-125.  
SEQ ID NO:51 is the determined cDNA sequence for LSC-134.  
10 SEQ ID NO:52 is the determined cDNA sequence for LSC-142.  
SEQ ID NO:53 is the determined cDNA sequence for LSC-144.  
SEQ ID NO:54 is the determined cDNA sequence for LSC-148.  
SEQ ID NO:55 is the determined cDNA sequence for LSC-149.  
SEQ ID NO:56 is the determined cDNA sequence for LSC-153.  
15 SEQ ID NO:57 is the determined cDNA sequence for LSC-163.  
SEQ ID NO:58 is the determined cDNA sequence for LSC-170.  
SEQ ID NO:59 is the determined cDNA sequence for LSC-171.  
SEQ ID NO:60 is the determined cDNA sequence for LSC-172.  
SEQ ID NO:61 is the determined cDNA sequence for LSC-175.  
20 SEQ ID NO:62 is the determined cDNA sequence for LSC-177.  
SEQ ID NO:63 is the determined cDNA sequence for LSC-182.  
SEQ ID NO:64 is the determined cDNA sequence for LSC-184.  
SEQ ID NO:65 is the determined cDNA sequence for LSC-189.  
SEQ ID NO:66 is the determined cDNA sequence for LSC-194.  
25 SEQ ID NO:67 is the determined cDNA sequence for LSC-195.  
SEQ ID NO:68 is the determined cDNA sequence for LSC-196.  
SEQ ID NO:69 is the determined cDNA sequence for LSC-199.  
SEQ ID NO:70 is the determined cDNA sequence for LSC-202.  
SEQ ID NO:71 is the determined cDNA sequence for LSC-203.

SEQ ID NO:72 is the determined cDNA sequence for LSC-205.  
SEQ ID NO:73 is the determined cDNA sequence for LSC-206.  
SEQ ID NO:74 is the determined cDNA sequence for LSC-210.  
SEQ ID NO:75 is the determined cDNA sequence for LSC-215.  
5 SEQ ID NO:76 is the determined cDNA sequence for LSC-218.  
SEQ ID NO:77 is the determined cDNA sequence for clone 48060.  
SEQ ID NO:78 is the determined cDNA sequence for clone 48069.  
SEQ ID NO:79 is the determined cDNA sequence for clone 48071.  
10 SEQ ID NO:80 is the determined cDNA sequence for clone 48080.  
SEQ ID NO:81 is the determined cDNA sequence for clone 48090.  
SEQ ID NO:82 is the determined cDNA sequence for clone 48102.  
SEQ ID NO:83 is the determined cDNA sequence for clone 48112.  
SEQ ID NO:84 is the determined cDNA sequence for clone 48118.  
SEQ ID NO:85 is the determined cDNA sequence for clone 48125.  
15 SEQ ID NO:86 is the determined cDNA sequence for clone 48129.  
SEQ ID NO:87 is the determined cDNA sequence for clone 48134.  
SEQ ID NO:88 is the determined cDNA sequence for clone 48135.  
SEQ ID NO:89 is the determined cDNA sequence for clone 48137.  
SEQ ID NO:90 is the determined cDNA sequence for clone 48138.  
20 SEQ ID NO:91 is the determined cDNA sequence for clone 48142.  
SEQ ID NO:92 is the determined cDNA sequence for clone 48143.  
SEQ ID NO:93 is the determined cDNA sequence for clone 48149.  
SEQ ID NO:94 is the determined cDNA sequence for clone 48150.  
SEQ ID NO:95 is the determined cDNA sequence for clone 48179.  
25 SEQ ID NO:96 is the determined cDNA sequence for clone 48183.  
SEQ ID NO:97 is the determined cDNA sequence for clone 48193.  
SEQ ID NO:98 is the determined cDNA sequence for clone 48196.  
SEQ ID NO:99 is the determined cDNA sequence for clone 48202.  
SEQ ID NO:100 is the determined cDNA sequence for clone 48204.

SEQ ID NO:101 is the determined cDNA sequence for clone 48205.  
SEQ ID NO:102 is the determined cDNA sequence for clone 48206.  
SEQ ID NO:103 is the determined cDNA sequence for clone 48211.  
SEQ ID NO:104 is the determined cDNA sequence for clone 48216.  
5 SEQ ID NO:105 is the determined cDNA sequence for clone 48219.  
SEQ ID NO:106 is the determined cDNA sequence for clone 48223.  
SEQ ID NO:107 is the determined cDNA sequence for clone 48224.  
SEQ ID NO:108 is the determined cDNA sequence for clone 48225.  
SEQ ID NO:109 is the determined cDNA sequence for clone 48228.  
10 SEQ ID NO:110 is the determined cDNA sequence for clone 48236.  
SEQ ID NO:111 is the determined cDNA sequence for clone lcl/15745.  
SEQ ID NO:112 is the determined cDNA sequence for clone lcl/16256.  
SEQ ID NO:113 is the determined cDNA sequence for clone lcl/21736.  
SEQ ID NO:114 is the determined cDNA sequence for clone lcl/22291.  
15 SEQ ID NO:115 is the determined cDNA sequence for clone lcl/24845.  
SEQ ID NO:116 is the determined cDNA sequence for clone lcl/24847.  
SEQ ID NO:117 is the determined cDNA sequence for clone lcl/24848.  
SEQ ID NO:118 is the determined cDNA sequence for clone lcl/24849.  
SEQ ID NO:119 is the determined cDNA sequence for clone lcl/24851.  
20 SEQ ID NO:120 is the determined cDNA sequence for clone lcl/24852.  
SEQ ID NO:121 is the determined cDNA sequence for clone lcl/24854.  
SEQ ID NO:122 is the determined cDNA sequence for clone lcl/24855.  
SEQ ID NO:123 is the determined cDNA sequence for clone lcl/24857.  
SEQ ID NO:124 is the determined cDNA sequence for clone lcl/24859.  
25 SEQ ID NO:125 is the determined cDNA sequence for clone lcl/24864.  
SEQ ID NO:126 is the determined cDNA sequence for clone lcl/24865.  
SEQ ID NO:127 is the determined cDNA sequence for clone lcl/24866.  
SEQ ID NO:128 is the determined cDNA sequence for clone lcl/24867.  
SEQ ID NO:129 is the determined cDNA sequence for clone lcl/24869.

SEQ ID NO:130 is the determined cDNA sequence for clone lcl/24871.  
SEQ ID NO:131 is the determined cDNA sequence for clone lcl/24873.  
SEQ ID NO:132 is the determined cDNA sequence for clone lcl/24874.  
SEQ ID NO:133 is the determined cDNA sequence for clone lcl/26008.  
5 SEQ ID NO:134 is the determined cDNA sequence for clone lcl/56871.  
SEQ ID NO:135 is the determined cDNA sequence for clone lcl/57480.  
SEQ ID NO:136 is the determined cDNA sequence for clone lcl/57499.  
SEQ ID NO:137 is the determined cDNA sequence for clone lcl/16785.  
SEQ ID NO:138 is the determined cDNA sequence for clone lcl/16787.  
10 SEQ ID NO:139 is the determined cDNA sequence for clone lcl/22175.  
SEQ ID NO:140 is the determined cDNA sequence for clone lcl/29484.  
SEQ ID NO:141 is the determined cDNA sequence for clone lcl/30354.  
SEQ ID NO:142 is the determined cDNA sequence for clone lcl/56868.  
SEQ ID NO:143 is the determined cDNA sequence for clone SCC2-1.  
15 SEQ ID NO:144 is the determined cDNA sequence for clone SCC2-2.  
SEQ ID NO:145 is the determined cDNA sequence for clone SCC2-4.  
SEQ ID NO:146 is the determined cDNA sequence for clone SCC2-5.  
SEQ ID NO:147 is the determined cDNA sequence for clone SCC2-7.  
SEQ ID NO:148 is the determined cDNA sequence for clone SCC2-9.  
20 SEQ ID NO:149 is the determined cDNA sequence for clone SCC2-10.  
SEQ ID NO:150 is the determined cDNA sequence for clone SCC2-11.  
SEQ ID NO:151 is the determined cDNA sequence for clone SCC2-12.  
SEQ ID NO:152 is the determined cDNA sequence for clone SCC2-13.  
SEQ ID NO:153 is the determined cDNA sequence for clone SCC2-14.  
25 SEQ ID NO:154 is the determined cDNA sequence for clone SCC2-17.  
SEQ ID NO:155 is the determined cDNA sequence for clone SCC2-18.  
SEQ ID NO:156 is the determined cDNA sequence for clone SCC2-20.  
SEQ ID NO:157 is the determined cDNA sequence for clone SCC2-23.  
SEQ ID NO:158 is the determined cDNA sequence for clone SCC2-24.

SEQ ID NO:159 is the determined cDNA sequence for clone SCC2-27.

SEQ ID NO:160 is the determined cDNA sequence for clone SCC2-29.

SEQ ID NO:161 is the determined cDNA sequence for clone SCC2-30.

SEQ ID NO:162 is the determined cDNA sequence for clone SCC2-31.

5       SEQ ID NO:163 is the determined cDNA sequence for clone SCC2-33.

SEQ ID NO:164 is the determined cDNA sequence for clone SCC2-35.

SEQ ID NO:165 is the determined cDNA sequence for clone SCC2-36.

SEQ ID NO:166 is the determined cDNA sequence for clone SCC2-37.

SEQ ID NO:167 is the determined cDNA sequence for clone SCC2-38.

10      SEQ ID NO:168 is the determined cDNA sequence for clone SCC2-39.

SEQ ID NO:169 is the determined cDNA sequence for clone SCC2-40.

SEQ ID NO:170 is the determined cDNA sequence for clone SCC2-43.

SEQ ID NO:171 is the determined cDNA sequence for clone SCC2-44.

SEQ ID NO:172 is the determined cDNA sequence for clone SCC2-46.

15      SEQ ID NO:173 is the determined cDNA sequence for clone SCC2-47.

SEQ ID NO:174 is the determined cDNA sequence for clone SCC2-48.

SEQ ID NO:175 is the determined cDNA sequence for clone SCC2-51.

SEQ ID NO:176 is the determined cDNA sequence for clone SCC2-52.

SEQ ID NO:177 is the determined cDNA sequence for clone SCC2-53.

20      SEQ ID NO:178 is the determined cDNA sequence for clone SCC2-54.

SEQ ID NO:179 is the determined cDNA sequence for clone SCC2-57.

SEQ ID NO:180 is the determined cDNA sequence for clone SCC2-58.

SEQ ID NO:181 is the determined cDNA sequence for clone SCC2-60.

SEQ ID NO:182 is the determined cDNA sequence for clone SCC2-64.

25      SEQ ID NO:183 is the determined cDNA sequence for clone SCC2-66.

SEQ ID NO:184 is the determined cDNA sequence for clone SCC2-68.

SEQ ID NO:185 is the determined cDNA sequence for clone SCC2-69.

SEQ ID NO:186 is the determined cDNA sequence for clone SCC2-70.

SEQ ID NO:187 is the determined cDNA sequence for clone SCC2-75.

SEQ ID NO:188 is the determined cDNA sequence for clone SCC2-77.  
SEQ ID NO:189 is the determined cDNA sequence for clone SCC2-78.  
SEQ ID NO:190 is the determined cDNA sequence for clone SCC2-79.  
SEQ ID NO:191 is the determined cDNA sequence for clone SCC2-80.  
5 SEQ ID NO:192 is the determined cDNA sequence for clone SCC2-84.  
SEQ ID NO:193 is the determined cDNA sequence for clone SCC2-85.  
SEQ ID NO:194 is the determined cDNA sequence for clone SCC2-91.  
SEQ ID NO:195 is the determined cDNA sequence for clone SCC2-92.  
SEQ ID NO:196 is the determined cDNA sequence for clone SCC2-95.  
10 SEQ ID NO:197 is the determined cDNA sequence for clone SCC2-96.  
SEQ ID NO:198 is the determined cDNA sequence for clone SCC2-97.  
SEQ ID NO:199 is the determined cDNA sequence for clone SCC2-98.  
SEQ ID NO:200 is the determined cDNA sequence for clone SCC2-100.  
SEQ ID NO:201 is the determined cDNA sequence for clone SCC2-101.  
15 SEQ ID NO:202 is the determined cDNA sequence for clone SCC2-102.  
SEQ ID NO:203 is the determined cDNA sequence for clone SCC2-103.  
SEQ ID NO:204 is the determined cDNA sequence for clone SCC2-104.  
SEQ ID NO:205 is the determined cDNA sequence for clone SCC2-107.  
SEQ ID NO:206 is the determined cDNA sequence for clone SCC2-108.  
20 SEQ ID NO:207 is the determined cDNA sequence for clone SCC2-110.  
SEQ ID NO:208 is the determined cDNA sequence for clone SCC2-112.  
SEQ ID NO:209 is the determined cDNA sequence for clone SCC2-116.  
SEQ ID NO:210 is the determined cDNA sequence for clone SCC2-124.  
SEQ ID NO:211 is the determined cDNA sequence for clone SCC2-125.  
25 SEQ ID NO:212 is the determined cDNA sequence for clone SCC2-131.  
SEQ ID NO:213 is the determined cDNA sequence for clone SCC2-137.  
SEQ ID NO:214 is the determined cDNA sequence for clone SCC2-143.  
SEQ ID NO:215 is the determined cDNA sequence for clone SCC2-146.  
SEQ ID NO:216 is the determined cDNA sequence for clone SCC2-154.

SEQ ID NO:217 is the determined cDNA sequence for clone SCC2-164.  
SEQ ID NO:218 is the determined cDNA sequence for clone SCC2-179.  
SEQ ID NO:219 is the determined cDNA sequence for clone SCC2-183.  
SEQ ID NO:220 is the determined cDNA sequence for clone SCC2-187.  
5 SEQ ID NO:221 is the determined cDNA sequence for clone SCC2-188.  
SEQ ID NO:222 is the determined cDNA sequence for clone SCC2-232.  
SEQ ID NO:223 is the determined cDNA sequence for clone SCC2-236.  
SEQ ID NO:224 is the determined cDNA sequence for clone SCC2-260.  
SEQ ID NO:225 is the determined cDNA sequence for clone SCC2-261.  
10 SEQ ID NO:226 is the determined cDNA sequence for clone SCC2-266.  
SEQ ID NO:227 is the determined cDNA sequence for clone SCC2-275.  
SEQ ID NO:228 is the determined cDNA sequence for clone SCC2-283.  
SEQ ID NO:229 is the determined cDNA extended sequence for clone  
SCC2-5, which relates to SEQ ID NO:146.  
15 SEQ ID NO:230 is the determined cDNA extended sequence for clone  
SCC2-14, which relates to SEQ ID NO:153.  
SEQ ID NO:231 is the determined cDNA sequence for clone SCC2-50.  
SEQ ID NO:232 is the determined cDNA extended sequence for clone  
SCC2-51, which relates to SEQ ID NO:175.  
20 SEQ ID NO:233 is the amino acid sequence encoded by SEQ ID NO:229.  
SEQ ID NO:234 is the amino acid sequence encoded by SEQ ID NO:230.  
SEQ ID NO:235 is the amino acid sequence encoded by SEQ ID NO:231.  
SEQ ID NO:236 is the amino acid sequence encoded by SEQ ID NO:232.  
SEQ ID NO:237 is GenBank Accession No. CAA58926  
25 SEQ ID NO:238 is GenBank Accession No. BAA91327  
SEQ ID NO:239 is GenBank Accession No. BAA22955  
SEQ ID NO:240 is GenBank Accession No. NP\_004258  
SEQ ID NO:241 is GenBank Accession No. AAF61208  
SEQ ID NO:242 is GenBank Accession No. CAA26370

SEQ ID NO:243 is the determined cDNA sequence for '56908.1  
SEQ ID NO:244 is the determined cDNA sequence for '56909.1  
SEQ ID NO:245 is the determined cDNA sequence for '56911.1  
SEQ ID NO:246 is GenBank Accession No. AK000700  
5 SEQ ID NO:247 is the determined cDNA sequence for '56912.1  
SEQ ID NO:248 is GenBank Accession No. AB006624  
SEQ ID NO:249 is the determined cDNA sequence for '56913.1  
SEQ ID NO:250 is GenBank Accession No. NM\_004267  
10 SEQ ID NO:251 is the determined cDNA sequence for '56916.1  
SEQ ID NO:252 is the determined cDNA sequence for '56917.1  
SEQ ID NO:253 is the determined cDNA sequence for '56921.1  
SEQ ID NO:254 is GenBank Accession No. AF216751  
SEQ ID NO:255 is the determined cDNA sequence for '56922.1  
15 SEQ ID NO:256 is GenBank Accession No. X02530  
SEQ ID NO:257 is the determined cDNA sequence for '56923.1  
SEQ ID NO:258 is the determined cDNA sequence for 54533.1  
SEQ ID NO:259 is the determined cDNA sequence for 54534.1  
SEQ ID NO:260 is the determined cDNA sequence for 54536.1  
SEQ ID NO:261 is the determined cDNA sequence for 54538.1  
20 SEQ ID NO:262 is the determined cDNA sequence for 54540.1  
SEQ ID NO:263 is the determined cDNA sequence for 55084.1  
SEQ ID NO:264 is the determined cDNA sequence for 55086.1  
SEQ ID NO:265 is the determined cDNA sequence for 54555.1  
SEQ ID NO:266 is the determined cDNA sequence for 54557.1  
25 SEQ ID NO:267 is the determined cDNA sequence for 54564.1  
SEQ ID NO:268 is the determined cDNA sequence for 55098.1  
SEQ ID NO:269 is the determined cDNA sequence for 55473.1  
SEQ ID NO:270 is the determined cDNA sequence for 55104.1  
SEQ ID NO:271 is the determined cDNA sequence for 55105.1

SEQ ID NO:272 is the determined cDNA sequence for 55107.1  
SEQ ID NO:273 is the determined cDNA sequence for 55108.1  
SEQ ID NO:274 is the determined cDNA sequence for 55114.1  
SEQ ID NO:275 is the determined cDNA sequence for 55477.1  
5 SEQ ID NO:276 is the determined cDNA sequence for 55482.1  
SEQ ID NO:277 is the determined cDNA sequence for 55483.1  
SEQ ID NO:278 is the determined cDNA sequence for 55485.1  
SEQ ID NO:279 is the determined cDNA sequence for 55487.1  
SEQ ID NO:280 is the determined cDNA sequence for 55488.1  
10 SEQ ID NO:281 is the determined cDNA sequence for 55087.1  
SEQ ID NO:282 is the determined cDNA sequence for 55089.1  
SEQ ID NO:283 is the determined cDNA sequence for 55092.1  
SEQ ID NO:284 is the determined cDNA sequence for 55093.1  
SEQ ID NO:285 is the determined cDNA sequence for 56926.1  
15 SEQ ID NO:286 is the determined cDNA sequence for 56930.1  
SEQ ID NO:287 is the determined cDNA sequence for 56944.1  
SEQ ID NO:288 is the determined cDNA sequence for 56945.1  
SEQ ID NO:289 is the determined cDNA sequence for 55490.1  
SEQ ID NO:290 is the determined cDNA sequence for 55495.1  
20 SEQ ID NO:291 is the determined cDNA sequence for 55504.1  
SEQ ID NO:292 is the determined cDNA sequence for 55506.1  
SEQ ID NO:293 is the determined cDNA sequence for 56480.1  
SEQ ID NO:294 is the determined cDNA sequence for 56482.1  
SEQ ID NO:295 is the determined cDNA sequence for 56484.1  
25 SEQ ID NO:296 is the determined cDNA sequence for 56487.1  
SEQ ID NO:297 is the determined cDNA sequence for 56488.1  
SEQ ID NO:298 is the determined cDNA sequence for 56490.1  
SEQ ID NO:299 is the determined cDNA sequence for 56493.1  
SEQ ID NO:300 is the determined cDNA sequence for 56494.1

SEQ ID NO:301 is the determined cDNA sequence for 56495.1  
SEQ ID NO:302 is the determined cDNA sequence for 56499.1  
SEQ ID NO:303 is the determined cDNA sequence for 56517.1  
SEQ ID NO:304 is the determined cDNA sequence for 56952.1  
5 SEQ ID NO:305 is the determined cDNA sequence for 56953.1  
SEQ ID NO:306 is the determined cDNA sequence for 56959.1  
SEQ ID NO:307 is the determined cDNA sequence for 57139.1  
SEQ ID NO:308 is the determined cDNA sequence for 57078.1  
SEQ ID NO:309 is the determined cDNA sequence for 57092.1  
10 SEQ ID NO:310 is the determined cDNA sequence for 57099.1  
SEQ ID NO:311 is the determined cDNA sequence for 57100.1  
SEQ ID NO:312 is the determined cDNA sequence for 57105.1  
SEQ ID NO:313 is the determined cDNA sequence for 57111.1  
SEQ ID NO:314 is the determined cDNA sequence for 57117.1  
15 SEQ ID NO:315 is the determined cDNA sequence for 57121.1  
SEQ ID NO:316 is the determined cDNA sequence for 57124.1  
SEQ ID NO:317 is the determined cDNA sequence for 57125.1  
SEQ ID NO:318 is the determined cDNA sequence for 54800.2  
SEQ ID NO:319 is the determined cDNA sequence for 54802.2  
20 SEQ ID NO:320 is the determined cDNA sequence for 54803.2  
SEQ ID NO:321 is the determined cDNA sequence for 54805.2  
SEQ ID NO:322 is the determined cDNA sequence for 54806.2  
SEQ ID NO:323 is the determined cDNA sequence for 54809.2  
SEQ ID NO:324 is the determined cDNA sequence for 54810.2  
25 SEQ ID NO:325 is the determined cDNA sequence for 54813.2  
SEQ ID NO:326 is the determined cDNA sequence for 54814.2  
SEQ ID NO:327 is the determined cDNA sequence for 54816.2  
SEQ ID NO:328 is the determined cDNA sequence for 54817.2  
SEQ ID NO:329 is the determined cDNA sequence for 54819.2

SEQ ID NO:330 is the determined cDNA sequence for 54821.2  
SEQ ID NO:331 is the determined cDNA sequence for 54823.2  
SEQ ID NO:332 is the determined cDNA sequence for 54824.2  
SEQ ID NO:333 is the determined cDNA sequence for 54825.2  
5 SEQ ID NO:334 is the determined cDNA sequence for 54826.2  
SEQ ID NO:335 is the determined cDNA sequence for 54827.2  
SEQ ID NO:336 is the determined cDNA sequence for 54829.2  
SEQ ID NO:337 is the determined cDNA sequence for 54830.2  
10 SEQ ID NO:338 is the determined cDNA sequence for 54832.2  
SEQ ID NO:339 is the determined cDNA sequence for 55800.2  
SEQ ID NO:340 is the determined cDNA sequence for 55801.2  
SEQ ID NO:341 is the determined cDNA sequence for 55803.2  
15 SEQ ID NO:342 is the determined cDNA sequence for 55804.2  
SEQ ID NO:343 is the determined cDNA sequence for 55805.2  
SEQ ID NO:344 is the determined cDNA sequence for 55806.2  
SEQ ID NO:345 is the determined cDNA sequence for 55808.2  
20 SEQ ID NO:346 is the determined cDNA sequence for 55810.2  
SEQ ID NO:347 is the determined cDNA sequence for 55811.2  
SEQ ID NO:348 is the determined cDNA sequence for 55812.2  
SEQ ID NO:349 is the determined cDNA sequence for 55814.2  
25 SEQ ID NO:350 is the determined cDNA sequence for 55816.2  
SEQ ID NO:351 is the determined cDNA sequence for 55817.2  
SEQ ID NO:352 is the determined cDNA sequence for 55819.2  
SEQ ID NO:353 is the determined cDNA sequence for 55820.2  
SEQ ID NO:354 is the determined cDNA sequence for 55823.2  
30 SEQ ID NO:355 is the determined cDNA sequence for 55824.2  
SEQ ID NO:356 is the determined cDNA sequence for 55826.2  
SEQ ID NO:357 is the determined cDNA sequence for 55828.2  
SEQ ID NO:358 is the determined cDNA sequence for 55829.2

SEQ ID NO:359 is the determined cDNA sequence for 55831.2  
SEQ ID NO:360 is the determined cDNA sequence for 55832.2  
SEQ ID NO:361 is the determined cDNA sequence for 55833.2  
SEQ ID NO:362 is the determined cDNA sequence for 55834.2  
5 SEQ ID NO:363 is the determined cDNA sequence for 55835.2  
SEQ ID NO:364 is the determined cDNA sequence for 55838.2  
SEQ ID NO:365 is a predicted extended cDNA sequence for clone 48137  
(L578S) having the isolated sequence of SEQ ID NO:89)

10 SEQ ID NO:366 is the predicted amino acid encoded by SEQ ID NO:365  
SEQ ID NO:367 is the determined cDNA sequence for 49949.5  
SEQ ID NO:368 is the determined cDNA sequence for 49952.1  
SEQ ID NO:369 is the determined cDNA sequence for 49956;contig 29  
SEQ ID NO:370 is the determined cDNA sequence for 49960.4  
15 SEQ ID NO:371 is the determined cDNA sequence for 49961;contig 21  
SEQ ID NO:372 is the determined cDNA sequence for 49962.4  
SEQ ID NO:373 is the determined cDNA sequence for 49962.5  
SEQ ID NO:374 is the determined cDNA sequence for 49965.1  
SEQ ID NO:375 is the determined cDNA sequence for 49966.1  
SEQ ID NO:376 is the determined cDNA sequence for 49971.1  
20 SEQ ID NO:377 is the determined cDNA sequence for 49975.1  
SEQ ID NO:378 is the determined cDNA sequence for 49982.1  
SEQ ID NO:379 is the determined cDNA sequence for 49986.1  
SEQ ID NO:380 is the determined cDNA sequence for 49988.1  
SEQ ID NO:381 is the determined cDNA sequence for 49993.1  
25 SEQ ID NO:382 is the determined cDNA sequence for 49995.1  
SEQ ID NO:383 is the determined cDNA sequence for 49996;contig 22  
SEQ ID NO:384 is the determined cDNA sequence for 49999.1  
SEQ ID NO:385 is the determined cDNA sequence for 50006;contig 23  
SEQ ID NO:386 is the determined cDNA sequence for 50007.1

SEQ ID NO:387 is the determined cDNA sequence for 50009.3  
SEQ ID NO:388 is the determined cDNA sequence for 50014.1  
SEQ ID NO:389 is the determined cDNA sequence for 50016;contig 24  
SEQ ID NO:390 is the determined cDNA sequence for 50017.1  
5 SEQ ID NO:391 is the determined cDNA sequence for 50019.1  
SEQ ID NO:392 is the determined cDNA sequence for 50022.1  
SEQ ID NO:393 is the determined cDNA sequence for 50023.1  
SEQ ID NO:394 is the determined cDNA sequence for 50024.1  
SEQ ID NO:395 is the determined cDNA sequence for 50033.1  
10 SEQ ID NO:396 is an extended cDNA sequence for SCC2-54 (SEQ ID  
NO:178)  
SEQ ID NO:397 is the amino acid sequence encoded by SEQ ID NO:396  
SEQ ID NO:398 is the determined cDNA sequence for 56908.1  
SEQ ID NO:399 is the determined cDNA sequence for 56911.1  
15 SEQ ID NO:400 is the determined cDNA sequence for 56912.1  
SEQ ID NO:401 is the determined cDNA sequence for 56913.1  
SEQ ID NO:402 is the determined cDNA sequence for 56916.1  
SEQ ID NO:403 is the determined cDNA sequence for 56917.1  
SEQ ID NO:404 is the determined cDNA sequence for 56921.1  
20 SEQ ID NO:405 is the determined cDNA sequence for 56922.1  
SEQ ID NO:406 is the determined cDNA sequence for 56923.1  
SEQ ID NO:407 is the determined cDNA sequence for 60974.1  
SEQ ID NO:408 is the determined cDNA sequence for 60976.1  
SEQ ID NO:409 is the determined cDNA sequence for 60977.1  
25 SEQ ID NO:410 is the determined cDNA sequence for 60978.1  
SEQ ID NO:411 is the determined cDNA sequence for 60980.1  
SEQ ID NO:412 is an extended cDNA sequence for LSC-49 (SEQ ID  
NO:29)  
SEQ ID NO:413 is the amino acid sequence encoded by SEQ ID NO:412

SEQ ID NO:414 is an extended cDNA sequence for LSC-39 (SEQ ID NO:26)

SEQ ID NO:415 is an extended cDNA sequence for LSC-46 (SEQ ID NO:28)

5 SEQ ID NO:416 is an extended cDNA sequence for LSC-49 (SEQ ID NO:29)

SEQ ID NO:417 is an extended cDNA sequence for LSC-51 (SEQ ID NO:30)

10 SEQ ID NO:418 is an extended cDNA sequence for LSC-55 (SEQ ID NO:32)

SEQ ID NO:419 is an extended cDNA sequence for LSC-64 (SEQ ID NO:35)

SEQ ID NO:420 is an extended cDNA sequence for LSC-78 (SEQ ID NO:42)

15 SEQ ID NO:421 is an extended cDNA sequence for LSC-103 (SEQ ID NO:47)

SEQ ID NO:422 is an extended cDNA sequence for LSC-144 (SEQ ID NO:53)

20 SEQ ID NO:423 is an extended cDNA sequence for LSC-148 (SEQ ID NO:54)

SEQ ID NO:424 is an extended cDNA sequence for LSC-210 (SEQ ID NO:74)

SEQ ID NO:425 is the amino acid sequence encoded by SEQ ID NO:414

SEQ ID NO:426 is the amino acid sequence encoded by SEQ ID NO:415

25 SEQ ID NO:427 is the amino acid sequence encoded by SEQ ID NO:416

SEQ ID NO:428 is the amino acid sequence encoded by SEQ ID NO:417

SEQ ID NO:429 is the amino acid sequence encoded by SEQ ID NO:418

SEQ ID NO:430 is the amino acid sequence encoded by SEQ ID NO:419

SEQ ID NO:431 is the amino acid sequence encoded by SEQ ID NO:420

SEQ ID NO:432 is the amino acid sequence encoded by SEQ ID NO:421  
SEQ ID NO:433 is the amino acid sequence encoded by SEQ ID NO:422  
SEQ ID NO:434 is the amino acid sequence encoded by SEQ ID NO:423  
SEQ ID NO:435 is the amino acid sequence encoded by SEQ ID NO:424  
5       SEQ ID NO:436 is the amino acid sequence encoded by a second open  
reading frame (ORF-2) of clone SCC2-51, SEQ ID NO:175  
SEQ ID NO:437 is the determined cDNA sequence for SCC2-16.  
SEQ ID NO:438 is the determined cDNA sequence for SCC2-28.  
SEQ ID NO:439 is the determined cDNA sequence for SCC2-62.  
10       SEQ ID NO:440 is the determined cDNA sequence for SCC3-90.

#### DETAILED DESCRIPTION OF THE INVENTION

The present invention is directed generally to compositions and their use in  
15 the therapy and diagnosis of cancer, particularly lung cancer. As described further below,  
illustrative compositions of the present invention include, but are not restricted to,  
polypeptides, particularly immunogenic polypeptides, polynucleotides encoding such  
polypeptides, antibodies and other binding agents, antigen presenting cells (APCs) and  
immune system cells (e.g., T cells).

20       The practice of the present invention will employ, unless indicated  
specifically to the contrary, conventional methods of virology, immunology, microbiology,  
molecular biology and recombinant DNA techniques within the skill of the art, many of  
which are described below for the purpose of illustration. Such techniques are explained  
fully in the literature. See, e.g., Sambrook, et al. Molecular Cloning: A Laboratory Manual  
25 (2nd Edition, 1989); Maniatis et al. Molecular Cloning: A Laboratory Manual (1982); DNA  
Cloning: A Practical Approach, vol. I & II (D. Glover, ed.); Oligonucleotide Synthesis (N.  
Gait, ed., 1984); Nucleic Acid Hybridization (B. Hames & S. Higgins, eds., 1985);

Transcription and Translation (B. Hames & S. Higgins, eds., 1984); Animal Cell Culture (R. Freshney, ed., 1986); Perbal, A Practical Guide to Molecular Cloning (1984).

All publications, patents and patent applications cited herein, whether supra or infra, are hereby incorporated by reference in their entirety.

5 As used in this specification and the appended claims, the singular forms "a," "an" and "the" include plural references unless the content clearly dictates otherwise.

### Polypeptide Compositions

As used herein, the term "polypeptide" is used in its conventional meaning, i.e., as a sequence of amino acids. The polypeptides are not limited to a specific length of 10 the product; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide, and such terms may be used interchangeably herein unless specifically indicated otherwise. This term also does not refer to or exclude post-expression modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like, as well as other modifications known in the art, both 15 naturally occurring and non-naturally occurring. A polypeptide may be an entire protein, or a subsequence thereof. Particular polypeptides of interest in the context of this invention are amino acid subsequences comprising epitopes, i.e., antigenic determinants substantially responsible for the immunogenic properties of a polypeptide and being capable of evoking an immune response.

20 Particularly illustrative polypeptides of the present invention comprise those encoded by a polynucleotide sequence set forth in any one of SEQ ID NOS: 1-232, 243-396, 398-412, 414-424 and 437-440, or a sequence that hybridizes under moderately stringent conditions, or, alternatively, under highly stringent conditions, to a polynucleotide sequence set forth in any one of SEQ ID NOS: 1-232, 243-396, 398-412, 414-424 and 437-25 440. Certain illustrative polypeptides of the invention comprise amino acid sequences as set forth in any one of SEQ ID NOS: 229-232, 237-242, 397, 413 and 425-436.

The polypeptides of the present invention are sometimes herein referred to as lung tumor proteins or lung tumor polypeptides, as an indication that their identification

has been based at least in part upon their increased levels of expression in lung tumor samples. Thus, a "lung tumor polypeptide" or "lung tumor protein," refers generally to a polypeptide sequence of the present invention, or a polynucleotide sequence encoding such a polypeptide, that is expressed in a substantial proportion of lung tumor samples, for 5 example preferably greater than about 20%, more preferably greater than about 30%, and most preferably greater than about 50% or more of lung tumor samples tested, at a level that is at least two fold, and preferably at least five fold, greater than the level of expression in normal tissues, as determined using a representative assay provided herein. A lung tumor polypeptide sequence of the invention, based upon its increased level of expression 10 in tumor cells, has particular utility both as a diagnostic marker as well as a therapeutic target, as further described below.

In certain preferred embodiments, the polypeptides of the invention are immunogenic, *i.e.*, they react detectably within an immunoassay (such as an ELISA or T-cell stimulation assay) with antisera and/or T-cells from a patient with lung cancer. 15 Screening for immunogenic activity can be performed using techniques well known to the skilled artisan. For example, such screens can be performed using methods such as those described in Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In one illustrative example, a polypeptide may be immobilized on a solid support and contacted with patient sera to allow binding of antibodies within the sera 20 to the immobilized polypeptide. Unbound sera may then be removed and bound antibodies detected using, for example,  $^{125}\text{I}$ -labeled Protein A.

As would be recognized by the skilled artisan, immunogenic portions of the polypeptides disclosed herein are also encompassed by the present invention. An "immunogenic portion," as used herein, is a fragment of an immunogenic polypeptide of 25 the invention that itself is immunologically reactive (*i.e.*, specifically binds) with the B-cells and/or T-cell surface antigen receptors that recognize the polypeptide. Immunogenic portions may generally be identified using well known techniques, such as those summarized in Paul, *Fundamental Immunology*, 3rd ed., 243-247 (Raven Press, 1993) and references cited therein. Such techniques include screening polypeptides for the ability to

react with antigen-specific antibodies, antisera and/or T-cell lines or clones. As used herein, antisera and antibodies are "antigen-specific" if they specifically bind to an antigen (*i.e.*, they react with the protein in an ELISA or other immunoassay, and do not react detectably with unrelated proteins). Such antisera and antibodies may be prepared as 5 described herein, and using well-known techniques.

In one preferred embodiment, an immunogenic portion of a polypeptide of the present invention is a portion that reacts with antisera and/or T-cells at a level that is not substantially less than the reactivity of the full-length polypeptide (*e.g.*, in an ELISA and/or T-cell reactivity assay). Preferably, the level of immunogenic activity of the immunogenic 10 portion is at least about 50%, preferably at least about 70% and most preferably greater than about 90% of the immunogenicity for the full-length polypeptide. In some instances, preferred immunogenic portions will be identified that have a level of immunogenic activity greater than that of the corresponding full-length polypeptide, *e.g.*, having greater than about 100% or 150% or more immunogenic activity.

15 In certain other embodiments, illustrative immunogenic portions may include peptides in which an N-terminal leader sequence and/or transmembrane domain have been deleted. Other illustrative immunogenic portions will contain a small N- and/or C-terminal deletion (*e.g.*, 1-30 amino acids, preferably 5-15 amino acids), relative to the mature protein.

20 In another embodiment, a polypeptide composition of the invention may also comprise one or more polypeptides that are immunologically reactive with T cells and/or antibodies generated against a polypeptide of the invention, particularly a polypeptide having an amino acid sequence disclosed herein, or to an immunogenic fragment or variant thereof.

25 In another embodiment of the invention, polypeptides are provided that comprise one or more polypeptides that are capable of eliciting T cells and/or antibodies that are immunologically reactive with one or more polypeptides described herein, or one or more polypeptides encoded by contiguous nucleic acid sequences contained in the polynucleotide sequences disclosed herein, or immunogenic fragments or variants thereof,

or to one or more nucleic acid sequences which hybridize to one or more of these sequences under conditions of moderate to high stringency.

The present invention, in another aspect, provides polypeptide fragments comprising at least about 5, 10, 15, 20, 25, 50, or 100 contiguous amino acids, or more, 5 including all intermediate lengths, of a polypeptide compositions set forth herein, such as those set forth in SEQ ID NOS: 229-232, 237-242, 397, 413 and 425-436, or those encoded by a polynucleotide sequence set forth in a sequence of SEQ ID NOS: 1-232, 243-396, 398-412, 414-424 and 437-440.

In another aspect, the present invention provides variants of the polypeptide 10 compositions described herein. Polypeptide variants generally encompassed by the present invention will typically exhibit at least about 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more identity (determined as described below), along its length, to a polypeptide sequences set forth herein.

In one preferred embodiment, the polypeptide fragments and variants 15 provide by the present invention are immunologically reactive with an antibody and/or T-cell that reacts with a full-length polypeptide specifically set for the herein.

In another preferred embodiment, the polypeptide fragments and variants provided by the present invention exhibit a level of immunogenic activity of at least about 20 50%, preferably at least about 70%, and most preferably at least about 90% or more of that exhibited by a full-length polypeptide sequence specifically set forth herein.

A polypeptide "variant," as the term is used herein, is a polypeptide that typically differs from a polypeptide specifically disclosed herein in one or more substitutions, deletions, additions and/or insertions. Such variants may be naturally occurring or may be synthetically generated, for example, by modifying one or more of the 25 above polypeptide sequences of the invention and evaluating their immunogenic activity as described herein and/or using any of a number of techniques well known in the art.

For example, certain illustrative variants of the polypeptides of the invention include those in which one or more portions, such as an N-terminal leader sequence or transmembrane domain, have been removed. Other illustrative variants include variants in

which a small portion (*e.g.*, 1-30 amino acids, preferably 5-15 amino acids) has been removed from the N- and/or C-terminal of the mature protein.

In many instances, a variant will contain conservative substitutions. A "conservative substitution" is one in which an amino acid is substituted for another amino acid that has similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydrophobic nature of the polypeptide to be substantially unchanged. As described above, modifications may be made in the structure of the polynucleotides and polypeptides of the present invention and still obtain a functional molecule that encodes a variant or derivative polypeptide with desirable characteristics, *e.g.*, with immunogenic characteristics. When it is desired to alter the amino acid sequence of a polypeptide to create an equivalent, or even an improved, immunogenic variant or portion of a polypeptide of the invention, one skilled in the art will typically change one or more of the codons of the encoding DNA sequence according to Table 1.

For example, certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites on substrate molecules. Since it is the interactive capacity and nature of a protein that defines that protein's biological functional activity, certain amino acid sequence substitutions can be made in a protein sequence, and, of course, its underlying DNA coding sequence, and nevertheless obtain a protein with like properties. It is thus contemplated that various changes may be made in the peptide sequences of the disclosed compositions, or corresponding DNA sequences which encode said peptides without appreciable loss of their biological utility or activity.

TABLE 1

Amino Acids		Codons						
Alanine	Ala	A	GCA	GCC	GCG	GCU		
Cysteine	Cys	C	UGC	UGU				
Aspartic acid	Asp	D	GAC	GAU				
Glutamic acid	Glu	E	GAA	GAG				
Phenylalanine	Phe	F	UUC	UUU				
Glycine	Gly	G	GGA	GGC	GGG	GGU		
Histidine	His	H	CAC	CAU				
Isoleucine	Ile	I	AUA	AUC	AUU			
Lysine	Lys	K	AAA	AAG				
Leucine	Leu	L	UUA	UUG	CUA	CUC	CUG	CUU
Methionine	Met	M	AUG					
Asparagine	Asn	N	AAC	AAU				
Proline	Pro	P	CCA	CCC	CCG	CCU		
Glutamine	Gln	Q	CAA	CAG				
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGU
Serine	Ser	S	AGC	AGU	UCA	UCC	UCG	UCU
Threonine	Thr	T	ACA	ACC	ACG	ACU		
Valine	Val	V	GUA	GUC	GUG	GUU		
Tryptophan	Trp	W	UGG					
Tyrosine	Tyr	Y	UAC	UAU				

In making such changes, the hydropathic index of amino acids may be  
5 considered. The importance of the hydropathic amino acid index in conferring interactive  
biologic function on a protein is generally understood in the art (Kyte and Doolittle, 1982,  
incorporated herein by reference). It is accepted that the relative hydropathic character of  
the amino acid contributes to the secondary structure of the resultant protein, which in turn  
defines the interaction of the protein with other molecules, for example, enzymes,  
10 substrates, receptors, DNA, antibodies, antigens, and the like. Each amino acid has been

assigned a hydropathic index on the basis of its hydrophobicity and charge characteristics (Kyte and Doolittle, 1982). These values are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (-0.4); threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-5 1.6); histidine (-3.2); glutamate (-3.5); glutamine (-3.5); aspartate (-3.5); asparagine (-3.5); lysine (-3.9); and arginine (-4.5).

It is known in the art that certain amino acids may be substituted by other amino acids having a similar hydropathic index or score and still result in a protein with similar biological activity, *i.e.* still obtain a biological functionally equivalent protein. In 10 making such changes, the substitution of amino acids whose hydropathic indices are within  $\pm 2$  is preferred, those within  $\pm 1$  are particularly preferred, and those within  $\pm 0.5$  are even more particularly preferred. It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity. U. S. Patent 4,554,101 (specifically incorporated herein by reference in its entirety), states that the greatest local 15 average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with a biological property of the protein.

As detailed in U. S. Patent 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0  $\pm$  1); glutamate (+3.0  $\pm$  1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); 20 threonine (-0.4); proline (-0.5  $\pm$  1); alanine (-0.5); histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); tryptophan (-3.4). It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent, and in particular, an immunologically equivalent protein. In such changes, the 25 substitution of amino acids whose hydrophilicity values are within  $\pm 2$  is preferred, those within  $\pm 1$  are particularly preferred, and those within  $\pm 0.5$  are even more particularly preferred.

As outlined above, amino acid substitutions are generally therefore based on the relative similarity of the amino acid side-chain substituents, for example, their

hydrophobicity, hydrophilicity, charge, size, and the like. Exemplary substitutions that take various of the foregoing characteristics into consideration are well known to those of skill in the art and include: arginine and lysine; glutamate and aspartate; serine and threonine; glutamine and asparagine; and valine, leucine and isoleucine.

5. In addition, any polynucleotide may be further modified to increase stability *in vivo*. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends; the use of phosphorothioate or 2' O-methyl rather than phosphodiester linkages in the backbone; and/or the inclusion of nontraditional bases such as inosine, queosine and wybutosine, as well as acetyl- methyl-, thio- and other 10 modified forms of adenine, cytidine, guanine, thymine and uridine.

Amino acid substitutions may further be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity and/or the amphipathic nature of the residues. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids 15 with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine and valine; glycine and alanine; asparagine and glutamine; and serine, threonine, phenylalanine and tyrosine. Other groups of amino acids that may represent conservative changes include: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; and (5) phe, tyr, trp, his. A variant may also, or 20 alternatively, contain nonconservative changes. In a preferred embodiment, variant polypeptides differ from a native sequence by substitution, deletion or addition of five amino acids or fewer. Variants may also (or alternatively) be modified by, for example, the deletion or addition of amino acids that have minimal influence on the immunogenicity, secondary structure and hydropathic nature of the polypeptide.

25. As noted above, polypeptides may comprise a signal (or leader) sequence at the N-terminal end of the protein, which co-translationally or post-translationally directs transfer of the protein. The polypeptide may also be conjugated to a linker or other sequence for ease of synthesis, purification or identification of the polypeptide (e.g., poly-

His), or to enhance binding of the polypeptide to a solid support. For example, a polypeptide may be conjugated to an immunoglobulin Fc region.

When comparing polypeptide sequences, two sequences are said to be "identical" if the sequence of amino acids in the two sequences is the same when aligned 5 for maximum correspondence, as described below. Comparisons between two sequences are typically performed by comparing the sequences over a comparison window to identify and compare local regions of sequence similarity. A "comparison window" as used herein, refers to a segment of at least about 20 contiguous positions, usually 30 to about 75, 40 to about 50, in which a sequence may be compared to a reference sequence of the same 10 number of contiguous positions after the two sequences are optimally aligned.

Optimal alignment of sequences for comparison may be conducted using the Megalign program in the Lasergene suite of bioinformatics software (DNASTAR, Inc., Madison, WI), using default parameters. This program embodies several alignment schemes described in the following references: Dayhoff, M.O. (1978) A model of 15 evolutionary change in proteins – Matrices for detecting distant relationships. In Dayhoff, M.O. (ed.) *Atlas of Protein Sequence and Structure*, National Biomedical Research Foundation, Washington DC Vol. 5, Suppl. 3, pp. 345-358; Hein J. (1990) Unified Approach to Alignment and Phylogenies pp. 626-645 *Methods in Enzymology* vol. 183, Academic Press, Inc., San Diego, CA; Higgins, D.G. and Sharp, P.M. (1989) *CABIOS* 5:151-153; Myers, E.W. and Muller W. (1988) *CABIOS* 4:11-17; Robinson, E.D. (1971) *Comb. Theor* 11:105; Santou, N. Nes, M. (1987) *Mol. Biol. Evol.* 4:406-425; Sneath, P.H.A. and Sokal, R.R. (1973) *Numerical Taxonomy – the Principles and Practice of Numerical Taxonomy*, Freeman Press, San Francisco, CA; Wilbur, W.J. and Lipman, D.J. 20 (1983) *Proc. Natl. Acad. Sci. USA* 80:726-730.

25 Alternatively, optimal alignment of sequences for comparison may be conducted by the local identity algorithm of Smith and Waterman (1981) *Add. APL. Math* 2:482, by the identity alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity methods of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85: 2444, by computerized implementations of these algorithms (GAP,

BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), or by inspection.

One preferred example of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al. (1977) *Nucl. Acids Res.* 25:3389-3402 and Altschul et al. (1990) *J. Mol. Biol.* 215:403-410, respectively. BLAST and BLAST 2.0 can be used, for example with the parameters described herein, to determine percent sequence identity for the polynucleotides and polypeptides of the invention. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. For amino acid sequences, a scoring matrix can be used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment.

In one preferred approach, the "percentage of sequence identity" is determined by comparing two optimally aligned sequences over a window of comparison of at least 20 positions, wherein the portion of the polypeptide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) of 20 percent or less, usually 5 to 15 percent, or 10 to 12 percent, as compared to the reference sequences (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the reference sequence (*i.e.*, the window size) and multiplying the results by 100 to yield the percentage of sequence identity.

Within other illustrative embodiments, a polypeptide may be a fusion polypeptide that comprises multiple polypeptides as described herein, or that comprises at

least one polypeptide as described herein and an unrelated sequence, such as a known tumor protein. A fusion partner may, for example, assist in providing T helper epitopes (an immunological fusion partner), preferably T helper epitopes recognized by humans, or may assist in expressing the protein (an expression enhancer) at higher yields than the native recombinant protein. Certain preferred fusion partners are both immunological and expression enhancing fusion partners. Other fusion partners may be selected so as to increase the solubility of the polypeptide or to enable the polypeptide to be targeted to desired intracellular compartments. Still further fusion partners include affinity tags, which facilitate purification of the polypeptide.

Fusion polypeptides may generally be prepared using standard techniques, including chemical conjugation. Preferably, a fusion polypeptide is expressed as a recombinant polypeptide, allowing the production of increased levels, relative to a non-fused polypeptide, in an expression system. Briefly, DNA sequences encoding the polypeptide components may be assembled separately, and ligated into an appropriate expression vector. The 3' end of the DNA sequence encoding one polypeptide component is ligated, with or without a peptide linker, to the 5' end of a DNA sequence encoding the second polypeptide component so that the reading frames of the sequences are in phase. This permits translation into a single fusion polypeptide that retains the biological activity of both component polypeptides.

A peptide linker sequence may be employed to separate the first and second polypeptide components by a distance sufficient to ensure that each polypeptide folds into its secondary and tertiary structures. Such a peptide linker sequence is incorporated into the fusion polypeptide using standard techniques well known in the art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to adopt a flexible extended conformation; (2) their inability to adopt a secondary structure that could interact with functional epitopes on the first and second polypeptides; and (3) the lack of hydrophobic or charged residues that might react with the polypeptide functional epitopes. Preferred peptide linker sequences contain Gly, Asn and Ser residues. Other near neutral amino acids, such as Thr and Ala may also be used in the linker sequence. Amino acid

sequences which may be usefully employed as linkers include those disclosed in Maratea et al., *Gene* 40:39-46, 1985; Murphy et al., *Proc. Natl. Acad. Sci. USA* 83:8258-8262, 1986; U.S. Patent No. 4,935,233 and U.S. Patent No. 4,751,180. The linker sequence may generally be from 1 to about 50 amino acids in length. Linker sequences are not required 5. when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric interference.

The ligated DNA sequences are operably linked to suitable transcriptional or translational regulatory elements. The regulatory elements responsible for expression of DNA are located only 5' to the DNA sequence encoding the first polypeptides. Similarly, 10 stop codons required to end translation and transcription termination signals are only present 3' to the DNA sequence encoding the second polypeptide.

The fusion polypeptide can comprise a polypeptide as described herein together with an unrelated immunogenic protein, such as an immunogenic protein capable of eliciting a recall response. Examples of such proteins include tetanus, tuberculosis and 15 hepatitis proteins (see, for example, Stoute et al. *New Engl. J. Med.*, 336:86-91, 1997).

In one preferred embodiment, the immunological fusion partner is derived from a *Mycobacterium* sp., such as a *Mycobacterium tuberculosis*-derived Ra12 fragment. Ra12 compositions and methods for their use in enhancing the expression and/or immunogenicity of heterologous polynucleotide/polypeptide sequences is described in U.S. 20 Patent Application 60/158,585, the disclosure of which is incorporated herein by reference in its entirety. Briefly, Ra12 refers to a polynucleotide region that is a subsequence of a *Mycobacterium tuberculosis* MTB32A nucleic acid. MTB32A is a serine protease of 32 KD molecular weight encoded by a gene in virulent and avirulent strains of *M. tuberculosis*. The nucleotide sequence and amino acid sequence of MTB32A have been 25 described (for example, U.S. Patent Application 60/158,585; see also, Skeiky et al., *Infection and Immun.* (1999) 67:3998-4007, incorporated herein by reference). C-terminal fragments of the MTB32A coding sequence express at high levels and remain as a soluble polypeptides throughout the purification process. Moreover, Ra12 may enhance the immunogenicity of heterologous immunogenic polypeptides with which it is fused. One

preferred Ra12 fusion polypeptide comprises a 14 KD C-terminal fragment corresponding to amino acid residues 192 to 323 of MTB32A. Other preferred Ra12 polynucleotides generally comprise at least about 15 consecutive nucleotides, at least about 30 nucleotides, at least about 60 nucleotides, at least about 100 nucleotides, at least about 200 nucleotides, 5 or at least about 300 nucleotides that encode a portion of a Ra12 polypeptide. Ra12 polynucleotides may comprise a native sequence (*i.e.*, an endogenous sequence that encodes a Ra12 polypeptide or a portion thereof) or may comprise a variant of such a sequence. Ra12 polynucleotide variants may contain one or more substitutions, additions, deletions and/or insertions such that the biological activity of the encoded fusion 10 polypeptide is not substantially diminished, relative to a fusion polypeptide comprising a native Ra12 polypeptide. Variants preferably exhibit at least about 70% identity, more preferably at least about 80% identity and most preferably at least about 90% identity to a polynucleotide sequence that encodes a native Ra12 polypeptide or a portion thereof.

Within other preferred embodiments, an immunological fusion partner is 15 derived from protein D, a surface protein of the gram-negative bacterium *Haemophilus influenza B* (WO 91/18926). Preferably, a protein D derivative comprises approximately the first third of the protein (*e.g.*, the first N-terminal 100-110 amino acids), and a protein D derivative may be lipidated. Within certain preferred embodiments, the first 109 residues of a Lipoprotein D fusion partner is included on the N-terminus to provide the polypeptide 20 with additional exogenous T-cell epitopes and to increase the expression level in *E. coli* (thus functioning as an expression enhancer). The lipid tail ensures optimal presentation of the antigen to antigen presenting cells. Other fusion partners include the non-structural protein from influenzae virus, NS1 (hemagglutinin). Typically, the N-terminal 81 amino acids are used, although different fragments that include T-helper epitopes may be used.

25 In another embodiment, the immunological fusion partner is the protein known as LYTA, or a portion thereof (preferably a C-terminal portion). LYTA is derived from *Streptococcus pneumoniae*, which synthesizes an N-acetyl-L-alanine amidase known as amidase LYTA (encoded by the LytA gene; *Gene* 43:265-292, 1986). LYTA is an autolysin that specifically degrades certain bonds in the peptidoglycan backbone. The C-

terminal domain of the LYTA protein is responsible for the affinity to the choline or to some choline analogues such as DEAE. This property has been exploited for the development of *E. coli* C-LYTA expressing plasmids useful for expression of fusion proteins. Purification of hybrid proteins containing the C-LYTA fragment at the amino 5 terminus has been described (see *Biotechnology* 10:795-798, 1992). Within a preferred embodiment, a repeat portion of LYTA may be incorporated into a fusion polypeptide. A repeat portion is found in the C-terminal region starting at residue 178. A particularly preferred repeat portion incorporates residues 188-305.

Yet another illustrative embodiment involves fusion polypeptides, and the 10 polynucleotides encoding them, wherein the fusion partner comprises a targeting signal capable of directing a polypeptide to the endosomal/lysosomal compartment, as described in U.S. Patent No. 5,633,234. An immunogenic polypeptide of the invention, when fused with this targeting signal, will associate more efficiently with MHC class II molecules and thereby provide enhanced *in vivo* stimulation of CD4<sup>+</sup> T-cells specific for the polypeptide.

15 Polypeptides of the invention are prepared using any of a variety of well known synthetic and/or recombinant techniques, the latter of which are further described below. Polypeptides, portions and other variants generally less than about 150 amino acids can be generated by synthetic means, using techniques well known to those of ordinary skill in the art. In one illustrative example, such polypeptides are synthesized using any of the 20 commercially available solid-phase techniques, such as the Merrifield solid-phase synthesis method, where amino acids are sequentially added to a growing amino acid chain. See Merrifield, *J. Am. Chem. Soc.* 85:2149-2146, 1963. Equipment for automated synthesis of polypeptides is commercially available from suppliers such as Perkin Elmer/Applied BioSystems Division (Foster City, CA), and may be operated according to the 25 manufacturer's instructions.

In general, polypeptide compositions (including fusion polypeptides) of the invention are isolated. An "isolated" polypeptide is one that is removed from its original environment. For example, a naturally-occurring protein or polypeptide is isolated if it is separated from some or all of the coexisting materials in the natural system. Preferably,

such polypeptides are also purified, *e.g.*, are at least about 90% pure, more preferably at least about 95% pure and most preferably at least about 99% pure.

### Polynucleotide Compositions

The present invention, in other aspects, provides polynucleotide compositions. The terms "DNA" and "polynucleotide" are used essentially interchangeably herein to refer to a DNA molecule that has been isolated free of total genomic DNA of a particular species. "Isolated," as used herein, means that a polynucleotide is substantially away from other coding sequences, and that the DNA molecule does not contain large portions of unrelated coding DNA, such as large chromosomal fragments or other functional genes or polypeptide coding regions. Of course, this refers to the DNA molecule as originally isolated, and does not exclude genes or coding regions later added to the segment by the hand of man.

As will be understood by those skilled in the art, the polynucleotide compositions of this invention can include genomic sequences, extra-genomic and plasmid-encoded sequences and smaller engineered gene segments that express, or may be adapted to express, proteins, polypeptides, peptides and the like. Such segments may be naturally isolated, or modified synthetically by the hand of man.

As will be also recognized by the skilled artisan, polynucleotides of the invention may be single-stranded (coding or antisense) or double-stranded, and may be DNA (genomic, cDNA or synthetic) or RNA molecules. RNA molecules may include HnRNA molecules, which contain introns and correspond to a DNA molecule in a one-to-one manner, and mRNA molecules, which do not contain introns. Additional coding or non-coding sequences may, but need not, be present within a polynucleotide of the present invention, and a polynucleotide may, but need not, be linked to other molecules and/or support materials.

Polynucleotides may comprise a native sequence (*i.e.*, an endogenous sequence that encodes a polypeptide/protein of the invention or a portion thereof) or may

comprise a sequence that encodes a variant or derivative, preferably and immunogenic variant or derivative, of such a sequence.

Therefore, according to another aspect of the present invention, polynucleotide compositions are provided that comprise some or all of a polynucleotide sequence set forth in any one of SEQ ID NOS: 1-232, 243-396, 398-412, 414-424 and 437-440, complements of a polynucleotide sequence set forth in any one of SEQ ID NOS: 1-232, 243-396, 398-412, 414-424 and 437-440, and degenerate variants of a polynucleotide sequence set forth in any one of SEQ ID NOS: 1-232, 243-396, 398-412, 414-424 and 437-440. In certain preferred embodiments, the polynucleotide sequences set forth herein encode immunogenic polypeptides, as described above.

In other related embodiments, the present invention provides polynucleotide variants having substantial identity to the sequences disclosed herein in SEQ ID NOS: 1-232, 243-396, 398-412, 414-424 and 437-440, for example those comprising at least 70% sequence identity, preferably at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% or higher, sequence identity compared to a polynucleotide sequence of this invention using the methods described herein, (e.g., BLAST analysis using standard parameters, as described below). One skilled in this art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning and the like.

Typically, polynucleotide variants will contain one or more substitutions, additions, deletions and/or insertions, preferably such that the immunogenicity of the polypeptide encoded by the variant polynucleotide is not substantially diminished relative to a polypeptide encoded by a polynucleotide sequence specifically set forth herein). The term "variants" should also be understood to encompass homologous genes of xenogenic origin.

In additional embodiments, the present invention provides polynucleotide fragments comprising various lengths of contiguous stretches of sequence identical to or complementary to one or more of the sequences disclosed herein. For example,

polynucleotides are provided by this invention that comprise at least about 10, 15, 20, 30, 40, 50, 75, 100, 150, 200, 300, 400, 500 or 1000 or more contiguous nucleotides of one or more of the sequences disclosed herein as well as all intermediate lengths there between. It will be readily understood that "intermediate lengths", in this context, means any length 5 between the quoted values, such as 16, 17, 18, 19, *etc.*; 21, 22, 23, *etc.*; 30, 31, 32, *etc.*; 50, 51, 52, 53, *etc.*; 100, 101, 102, 103, *etc.*; 150, 151, 152, 153, *etc.*; including all integers through 200-500; 500-1,000, and the like.

In another embodiment of the invention, polynucleotide compositions are provided that are capable of hybridizing under moderate to high stringency conditions to a 10 polynucleotide sequence provided herein, or a fragment thereof, or a complementary sequence thereof. Hybridization techniques are well known in the art of molecular biology. For purposes of illustration, suitable moderately stringent conditions for testing the hybridization of a polynucleotide of this invention with other polynucleotides include 15 prewashing in a solution of 5 X SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0); hybridizing at 50°C-60°C, 5 X SSC, overnight; followed by washing twice at 65°C for 20 minutes with each of 2X, 0.5X and 0.2X SSC containing 0.1% SDS. One skilled in the art will understand that the stringency of hybridization can be readily manipulated, such as by altering the salt content of the hybridization solution and/or the temperature at which the 20 hybridization is performed. For example, in another embodiment, suitable highly stringent hybridization conditions include those described above, with the exception that the temperature of hybridization is increased, *e.g.*, to 60-65°C or 65-70°C.

In certain preferred embodiments, the polynucleotides described above, *e.g.*, 25 polynucleotide variants, fragments and hybridizing sequences, encode polypeptides that are immunologically cross-reactive with a polypeptide sequence specifically set forth herein. In other preferred embodiments, such polynucleotides encode polypeptides that have a level of immunogenic activity of at least about 50%, preferably at least about 70%, and more preferably at least about 90% of that for a polypeptide sequence specifically set forth herein.

The polynucleotides of the present invention, or fragments thereof, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length 5 may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol. For example, illustrative polynucleotide segments with total lengths of about 10,000, about 5000, about 3000, about 2,000, about 1,000, about 500, about 200, about 100, about 50 base pairs in length, and the 10 like, (including all intermediate lengths) are contemplated to be useful in many implementations of this invention.

When comparing polynucleotide sequences, two sequences are said to be "identical" if the sequence of nucleotides in the two sequences is the same when aligned for maximum correspondence, as described below. Comparisons between two sequences are 15 typically performed by comparing the sequences over a comparison window to identify and compare local regions of sequence similarity. A "comparison window" as used herein, refers to a segment of at least about 20 contiguous positions, usually 30 to about 75, 40 to about 50, in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned.

20 Optimal alignment of sequences for comparison may be conducted using the Megalign program in the Lasergene suite of bioinformatics software (DNASTAR, Inc., Madison, WI), using default parameters. This program embodies several alignment schemes described in the following references: Dayhoff, M.O. (1978) A model of evolutionary change in proteins – Matrices for detecting distant relationships. In Dayhoff, 25 M.O. (ed.) *Atlas of Protein Sequence and Structure*, National Biomedical Research Foundation, Washington DC Vol. 5, Suppl. 3, pp. 345-358; Hein J. (1990) Unified Approach to Alignment and Phylogenies pp. 626-645 *Methods in Enzymology* vol. 183, Academic Press, Inc., San Diego, CA; Higgins, D.G. and Sharp, P.M. (1989) *CABIOS* 5:151-153; Myers, E.W. and Muller W. (1988) *CABIOS* 4:11-17; Robinson, E.D. (1971)

*Comb. Theor* 11:105; Santou, N. Nes, M. (1987) *Mol. Biol. Evol.* 4:406-425; Sneath, P.H.A. and Sokal, R.R. (1973) *Numerical Taxonomy – the Principles and Practice of Numerical Taxonomy*, Freeman Press, San Francisco, CA; Wilbur, W.J. and Lipman, D.J. (1983) *Proc. Natl. Acad. Sci. USA* 80:726-730.

5           Alternatively, optimal alignment of sequences for comparison may be conducted by the local identity algorithm of Smith and Waterman (1981) *Add. APL. Math* 2:482, by the identity alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity methods of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85: 2444, by computerized implementations of these algorithms (GAP, 10 BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), or by inspection.

One preferred example of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al. (1977) *Nucl. Acids Res.* 25:3389-3402 15 and Altschul et al. (1990) *J. Mol. Biol.* 215:403-410, respectively. BLAST and BLAST 2.0 can be used, for example with the parameters described herein, to determine percent sequence identity for the polynucleotides of the invention. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. In one illustrative example, cumulative scores can be calculated using, for 20 nucleotide sequences, the parameters M (reward score for a pair of matching residues; always  $>0$ ) and N (penalty score for mismatching residues; always  $<0$ ). Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of 25 either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, and expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915) alignments, (B) of 50, expectation (E) of 10, M=5, N=-4 and a comparison of both strands.

Preferably, the "percentage of sequence identity" is determined by comparing two optimally aligned sequences over a window of comparison of at least 20 positions, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) of 20 percent or less, usually 5 to 15 percent, or 10 to 12 percent, as compared to the reference sequences (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid bases occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the reference sequence (*i.e.*, the window size) and multiplying the results by 100 to yield the percentage of sequence identity.

It will be appreciated by those of ordinary skill in the art that, as a result of the degeneracy of the genetic code, there are many nucleotide sequences that encode a polypeptide as described herein. Some of these polynucleotides bear minimal homology to the nucleotide sequence of any native gene. Nonetheless, polynucleotides that vary due to differences in codon usage are specifically contemplated by the present invention. Further, alleles of the genes comprising the polynucleotide sequences provided herein are within the scope of the present invention. Alleles are endogenous genes that are altered as a result of one or more mutations, such as deletions, additions and/or substitutions of nucleotides. The resulting mRNA and protein may, but need not, have an altered structure or function. Alleles may be identified using standard techniques (such as hybridization, amplification and/or database sequence comparison).

Therefore, in another embodiment of the invention, a mutagenesis approach, such as site-specific mutagenesis, is employed for the preparation of immunogenic variants and/or derivatives of the polypeptides described herein. By this approach, specific modifications in a polypeptide sequence can be made through mutagenesis of the underlying polynucleotides that encode them. These techniques provides a straightforward approach to prepare and test sequence variants, for example, incorporating one or more of

the foregoing considerations, by introducing one or more nucleotide sequence changes into the polynucleotide.

Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the deletion junction being traversed. Mutations may be employed in a selected polynucleotide sequence to improve, alter, decrease, modify, or otherwise change the properties of the polynucleotide itself, and/or alter the properties, activity, composition, 10 stability, or primary sequence of the encoded polypeptide.

In certain embodiments of the present invention, the inventors contemplate the mutagenesis of the disclosed polynucleotide sequences to alter one or more properties of the encoded polypeptide, such as the immunogenicity of a polypeptide vaccine. The techniques of site-specific mutagenesis are well-known in the art, and are widely used to 15 create variants of both polypeptides and polynucleotides. For example, site-specific mutagenesis is often used to alter a specific portion of a DNA molecule. In such embodiments, a primer comprising typically about 14 to about 25 nucleotides or so in length is employed, with about 5 to about 10 residues on both sides of the junction of the sequence being altered.

20 As will be appreciated by those of skill in the art, site-specific mutagenesis techniques have often employed a phage vector that exists in both a single stranded and double stranded form. Typical vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage are readily commercially-available and their use is generally well-known to those skilled in the art. Double-stranded plasmids are also 25 routinely employed in site directed mutagenesis that eliminates the step of transferring the gene of interest from a plasmid to a phage.

In general, site-directed mutagenesis in accordance herewith is performed by first obtaining a single-stranded vector or melting apart of two strands of a double-stranded vector that includes within its sequence a DNA sequence that encodes the desired peptide.

An oligonucleotide primer bearing the desired mutated sequence is prepared, generally synthetically. This primer is then annealed with the single-stranded vector, and subjected to DNA polymerizing enzymes such as *E. coli* polymerase I Klenow fragment, in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed 5 wherein one strand encodes the original non-mutated sequence and the second strand bears the desired mutation. This heteroduplex vector is then used to transform appropriate cells, such as *E. coli* cells, and clones are selected which include recombinant vectors bearing the mutated sequence arrangement.

The preparation of sequence variants of the selected peptide-encoding DNA 10 segments using site-directed mutagenesis provides a means of producing potentially useful species and is not meant to be limiting as there are other ways in which sequence variants of peptides and the DNA sequences encoding them may be obtained. For example, recombinant vectors encoding the desired peptide sequence may be treated with mutagenic agents, such as hydroxylamine, to obtain sequence variants. Specific details regarding 15 these methods and protocols are found in the teachings of Maloy *et al.*, 1994; Segal, 1976; Prokop and Bajpai, 1991; Kuby, 1994; and Maniatis *et al.*, 1982, each incorporated herein by reference, for that purpose.

As used herein, the term "oligonucleotide directed mutagenesis procedure" 20 refers to template-dependent processes and vector-mediated propagation which result in an increase in the concentration of a specific nucleic acid molecule relative to its initial concentration, or in an increase in the concentration of a detectable signal, such as amplification. As used herein, the term "oligonucleotide directed mutagenesis procedure" 25 is intended to refer to a process that involves the template-dependent extension of a primer molecule. The term template dependent process refers to nucleic acid synthesis of an RNA or a DNA molecule wherein the sequence of the newly synthesized strand of nucleic acid is dictated by the well-known rules of complementary base pairing (see, for example, Watson, 1987). Typically, vector mediated methodologies involve the introduction of the nucleic acid fragment into a DNA or RNA vector, the clonal amplification of the vector, and the recovery of the amplified nucleic acid fragment. Examples of such methodologies are

provided by U. S. Patent No. 4,237,224, specifically incorporated herein by reference in its entirety.

In another approach for the production of polypeptide variants of the present invention, recursive sequence recombination, as described in U.S. Patent No. 5,837,458, 5 may be employed. In this approach, iterative cycles of recombination and screening or selection are performed to "evolve" individual polynucleotide variants of the invention having, for example, enhanced immunogenic activity.

In other embodiments of the present invention, the polynucleotide sequences provided herein can be advantageously used as probes or primers for nucleic acid 10 hybridization. As such, it is contemplated that nucleic acid segments that comprise a sequence region of at least about 15 nucleotide long contiguous sequence that has the same sequence as, or is complementary to, a 15 nucleotide long contiguous sequence disclosed herein will find particular utility. Longer contiguous identical or complementary sequences, *e.g.*, those of about 20, 30, 40, 50, 100, 200, 500, 1000 (including all 15 intermediate lengths) and even up to full length sequences will also be of use in certain embodiments.

The ability of such nucleic acid probes to specifically hybridize to a sequence of interest will enable them to be of use in detecting the presence of complementary sequences in a given sample. However, other uses are also envisioned, 20 such as the use of the sequence information for the preparation of mutant species primers, or primers for use in preparing other genetic constructions.

Polynucleotide molecules having sequence regions consisting of contiguous nucleotide stretches of 10-14, 15-20, 30, 50, or even of 100-200 nucleotides or so (including intermediate lengths as well), identical or complementary to a polynucleotide 25 sequence disclosed herein, are particularly contemplated as hybridization probes for use in, *e.g.*, Southern and Northern blotting. This would allow a gene product, or fragment thereof, to be analyzed, both in diverse cell types and also in various bacterial cells. The total size of fragment, as well as the size of the complementary stretch(es), will ultimately depend on the intended use or application of the particular nucleic acid segment. Smaller

fragments will generally find use in hybridization embodiments, wherein the length of the contiguous complementary region may be varied, such as between about 15 and about 100 nucleotides, but larger contiguous complementarity stretches may be used, according to the length complementary sequences one wishes to detect.

5        The use of a hybridization probe of about 15-25 nucleotides in length allows the formation of a duplex molecule that is both stable and selective. Molecules having contiguous complementary sequences over stretches greater than 15 bases in length are generally preferred, though, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of specific hybrid molecules obtained. One will 10 generally prefer to design nucleic acid molecules having gene-complementary stretches of 15 to 25 contiguous nucleotides, or even longer where desired.

Hybridization probes may be selected from any portion of any of the sequences disclosed herein. All that is required is to review the sequences set forth herein, or to any continuous portion of the sequences, from about 15-25 nucleotides in length up to 15 and including the full length sequence, that one wishes to utilize as a probe or primer. The choice of probe and primer sequences may be governed by various factors. For example, one may wish to employ primers from towards the termini of the total sequence.

Small polynucleotide segments or fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical means, as is commonly practiced 20 using an automated oligonucleotide synthesizer. Also, fragments may be obtained by application of nucleic acid reproduction technology, such as the PCR™ technology of U. S. Patent 4,683,202 (incorporated herein by reference), by introducing selected sequences into recombinant vectors for recombinant production, and by other recombinant DNA techniques generally known to those of skill in the art of molecular biology.

25        The nucleotide sequences of the invention may be used for their ability to selectively form duplex molecules with complementary stretches of the entire gene or gene fragments of interest. Depending on the application envisioned, one will typically desire to employ varying conditions of hybridization to achieve varying degrees of selectivity of probe towards target sequence. For applications requiring high selectivity, one will

typically desire to employ relatively stringent conditions to form the hybrids, e.g., one will select relatively low salt and/or high temperature conditions, such as provided by a salt concentration of from about 0.02 M to about 0.15 M salt at temperatures of from about 50°C to about 70°C. Such selective conditions tolerate little, if any, mismatch between the 5 probe and the template or target strand, and would be particularly suitable for isolating related sequences.

Of course, for some applications, for example, where one desires to prepare mutants employing a mutant primer strand hybridized to an underlying template, less stringent (reduced stringency) hybridization conditions will typically be needed in order to 10 allow formation of the heteroduplex. In these circumstances, one may desire to employ salt conditions such as those of from about 0.15 M to about 0.9 M salt, at temperatures ranging from about 20°C to about 55°C. Cross-hybridizing species can thereby be readily identified as positively hybridizing signals with respect to control hybridizations. In any case, it is generally appreciated that conditions can be rendered more stringent by the addition of 15 increasing amounts of formamide, which serves to destabilize the hybrid duplex in the same manner as increased temperature. Thus, hybridization conditions can be readily manipulated, and thus will generally be a method of choice depending on the desired results.

According to another embodiment of the present invention, polynucleotide 20 compositions comprising antisense oligonucleotides are provided. Antisense oligonucleotides have been demonstrated to be effective and targeted inhibitors of protein synthesis, and, consequently, provide a therapeutic approach by which a disease can be treated by inhibiting the synthesis of proteins that contribute to the disease. The efficacy of antisense oligonucleotides for inhibiting protein synthesis is well established. For example, 25 the synthesis of polygalacturonase and the muscarine type 2 acetylcholine receptor are inhibited by antisense oligonucleotides directed to their respective mRNA sequences (U. S. Patent 5,739,119 and U. S. Patent 5,759,829). Further, examples of antisense inhibition have been demonstrated with the nuclear protein cyclin, the multiple drug resistance gene (MDG1), ICAM-1, E-selectin, STK-1, striatal GABA<sub>A</sub> receptor and human EGF (Jaskulski

*et al.*, *Science*. 1988 Jun 10;240(4858):1544-6; *Vasanthakumar and Ahmed, Cancer Commun.* 1989;1(4):225-32; *Peris et al., Brain Res Mol Brain Res.* 1998 Jun 15;57(2):310-20; U. S. Patent 5,801,154; U.S. Patent 5,789,573; U. S. Patent 5,718,709 and U.S. Patent 5,610,288). Antisense constructs have also been described that inhibit and can be used to 5 treat a variety of abnormal cellular proliferations, *e.g.* cancer (U. S. Patent 5,747,470; U. S. Patent 5,591,317 and U. S. Patent 5,783,683).

Therefore, in certain embodiments, the present invention provides oligonucleotide sequences that comprise all, or a portion of, any sequence that is capable of specifically binding to polynucleotide sequence described herein, or a complement thereof. 10 In one embodiment, the antisense oligonucleotides comprise DNA or derivatives thereof. In another embodiment, the oligonucleotides comprise RNA or derivatives thereof. In a third embodiment, the oligonucleotides are modified DNAs comprising a phosphorothioated modified backbone. In a fourth embodiment, the oligonucleotide sequences comprise peptide nucleic acids or derivatives thereof. In each case, preferred 15 compositions comprise a sequence region that is complementary, and more preferably substantially-complementary, and even more preferably, completely complementary to one or more portions of polynucleotides disclosed herein. Selection of antisense compositions specific for a given gene sequence is based upon analysis of the chosen target sequence and determination of secondary structure,  $T_m$ , binding energy, and relative stability. Antisense 20 compositions may be selected based upon their relative inability to form dimers, hairpins, or other secondary structures that would reduce or prohibit specific binding to the target mRNA in a host cell. Highly preferred target regions of the mRNA, are those which are at or near the AUG translation initiation codon, and those sequences which are substantially complementary to 5' regions of the mRNA. These secondary structure analyses and target 25 site selection considerations can be performed, for example, using v.4 of the OLIGO primer analysis software and/or the BLASTN 2.0.5 algorithm software (Altschul *et al.*, *Nucleic Acids Res.* 1997, 25(17):3389-402).

The use of an antisense delivery method employing a short peptide vector, termed MPG (27 residues), is also contemplated. The MPG peptide contains a

hydrophobic domain derived from the fusion sequence of HIV gp41 and a hydrophilic domain from the nuclear localization sequence of SV40 T-antigen (Morris *et al.*, Nucleic Acids Res. 1997 Jul 15;25(14):2730-6). It has been demonstrated that several molecules of the MPG peptide coat the antisense oligonucleotides and can be delivered into cultured 5 mammalian cells in less than 1 hour with relatively high efficiency (90%). Further, the interaction with MPG strongly increases both the stability of the oligonucleotide to nuclease and the ability to cross the plasma membrane.

According to another embodiment of the invention, the polynucleotide compositions described herein are used in the design and preparation of ribozyme 10 molecules for inhibiting expression of the tumor polypeptides and proteins of the present invention in tumor cells. Ribozymes are RNA-protein complexes that cleave nucleic acids in a site-specific fashion. Ribozymes have specific catalytic domains that possess endonuclease activity (Kim and Cech, Proc Natl Acad Sci U S A. 1987 Dec;84(24):8788-92; Forster and Symons, Cell. 1987 Apr 24;49(2):211-20). For example, a large number of 15 ribozymes accelerate phosphoester transfer reactions with a high degree of specificity, often cleaving only one of several phosphoesters in an oligonucleotide substrate (Cech *et al.*, Cell. 1981 Dec;27(3 Pt 2):487-96; Michel and Westhof, J Mol Biol. 1990 Dec 5;216(3):585-610; Reinhold-Hurek and Shub, Nature. 1992 May 14;357(6374):173-6). This specificity has been attributed to the requirement that the substrate bind via specific 20 base-pairing interactions to the internal guide sequence ("IGS") of the ribozyme prior to chemical reaction.

Six basic varieties of naturally-occurring enzymatic RNAs are known presently. Each can catalyze the hydrolysis of RNA phosphodiester bonds *in trans* (and thus can cleave other RNA molecules) under physiological conditions. In general, 25 enzymatic nucleic acids act by first binding to a target RNA. Such binding occurs through the target binding portion of a enzymatic nucleic acid which is held in close proximity to an enzymatic portion of the molecule that acts to cleave the target RNA. Thus, the enzymatic nucleic acid first recognizes and then binds a target RNA through complementary base-pairing, and once bound to the correct site, acts enzymatically to cut the target RNA.

Strategic cleavage of such a target RNA will destroy its ability to direct synthesis of an encoded protein. After an enzymatic nucleic acid has bound and cleaved its RNA target, it is released from that RNA to search for another target and can repeatedly bind and cleave new targets.

5        The enzymatic nature of a ribozyme is advantageous over many technologies, such as antisense technology (where a nucleic acid molecule simply binds to a nucleic acid target to block its translation) since the concentration of ribozyme necessary to affect a therapeutic treatment is lower than that of an antisense oligonucleotide. This advantage reflects the ability of the ribozyme to act enzymatically. Thus, a single ribozyme  
10 molecule is able to cleave many molecules of target RNA. In addition, the ribozyme is a highly specific inhibitor, with the specificity of inhibition depending not only on the base pairing mechanism of binding to the target RNA, but also on the mechanism of target RNA cleavage. Single mismatches, or base-substitutions, near the site of cleavage can completely eliminate catalytic activity of a ribozyme. Similar mismatches in antisense  
15 molecules do not prevent their action (Woolf *et al.*, Proc Natl Acad Sci U S A. 1992 Aug 15;89(16):7305-9). Thus, the specificity of action of a ribozyme is greater than that of an antisense oligonucleotide binding the same RNA site.

20        The enzymatic nucleic acid molecule may be formed in a hammerhead, hairpin, a hepatitis  $\delta$  virus, group I intron or RNaseP RNA (in association with an RNA guide sequence) or Neurospora VS RNA motif. Examples of hammerhead motifs are described by Rossi *et al.* Nucleic Acids Res. 1992 Sep 11;20(17):4559-65. Examples of hairpin motifs are described by Hampel *et al.* (Eur. Pat. Appl. Publ. No. EP 0360257), Hampel and Tritz, Biochemistry 1989 Jun 13;28(12):4929-33; Hampel *et al.*, Nucleic Acids Res. 1990 Jan 25;18(2):299-304 and U. S. Patent 5,631,359. An example of the  
25 hepatitis  $\delta$  virus motif is described by Perrotta and Been, Biochemistry. 1992 Dec 1;31(47):11843-52; an example of the RNaseP motif is described by Guerrier-Takada *et al.*, Cell. 1983 Dec;35(3 Pt 2):849-57; Neurospora VS RNA ribozyme motif is described by Collins (Saville and Collins, Cell. 1990 May 18;61(4):685-96; Saville and Collins, Proc Natl Acad Sci U S A. 1991 Oct 1;88(19):8826-30; Collins and Olive, Biochemistry. 1993

Mar 23;32(11):2795-9); and an example of the Group I intron is described in (U. S. Patent 4,987,071). All that is important in an enzymatic nucleic acid molecule of this invention is that it has a specific substrate binding site which is complementary to one or more of the target gene RNA regions, and that it have nucleotide sequences within or surrounding that 5 substrate binding site which impart an RNA cleaving activity to the molecule. Thus the ribozyme constructs need not be limited to specific motifs mentioned herein.

Ribozymes may be designed as described in Int. Pat. Appl. Publ. No. WO 93/23569 and Int. Pat. Appl. Publ. No. WO 94/02595, each specifically incorporated herein by reference) and synthesized to be tested *in vitro* and *in vivo*, as described. Such 10 ribozymes can also be optimized for delivery. While specific examples are provided, those in the art will recognize that equivalent RNA targets in other species can be utilized when necessary.

Ribozyme activity can be optimized by altering the length of the ribozyme binding arms, or chemically synthesizing ribozymes with modifications that prevent their 15 degradation by serum ribonucleases (see e.g., Int. Pat. Appl. Publ. No. WO 92/07065; Int. Pat. Appl. Publ. No. WO 93/15187; Int. Pat. Appl. Publ. No. WO 91/03162; Eur. Pat. Appl. Publ. No. 92110298.4; U. S. Patent 5,334,711; and Int. Pat. Appl. Publ. No. WO 94/13688, which describe various chemical modifications that can be made to the sugar moieties of enzymatic RNA molecules), modifications which enhance their efficacy in cells, and 20 removal of stem II bases to shorten RNA synthesis times and reduce chemical requirements.

Sullivan *et al.* (Int. Pat. Appl. Publ. No. WO 94/02595) describes the general methods for delivery of enzymatic RNA molecules. Ribozymes may be administered to cells by a variety of methods known to those familiar to the art, including, 25 but not restricted to, encapsulation in liposomes, by iontophoresis, or by incorporation into other vehicles, such as hydrogels, cyclodextrins, biodegradable nanocapsules, and bioadhesive microspheres. For some indications, ribozymes may be directly delivered *ex vivo* to cells or tissues with or without the aforementioned vehicles. Alternatively, the RNA/vehicle combination may be locally delivered by direct inhalation, by direct injection

or by use of a catheter, infusion pump or stent. Other routes of delivery include, but are not limited to, intravascular, intramuscular, subcutaneous or joint injection, aerosol inhalation, oral (tablet or pill form), topical, systemic, ocular, intraperitoneal and/or intrathecal delivery. More detailed descriptions of ribozyme delivery and administration are provided  
5 in Int. Pat. Appl. Publ. No. WO 94/02595 and Int. Pat. Appl. Publ. No. WO 93/23569, each specifically incorporated herein by reference.

Another means of accumulating high concentrations of a ribozyme(s) within cells is to incorporate the ribozyme-encoding sequences into a DNA expression vector. Transcription of the ribozyme sequences are driven from a promoter for eukaryotic RNA  
10 polymerase I (pol I), RNA polymerase II (pol II), or RNA polymerase III (pol III). Transcripts from pol II or pol III promoters will be expressed at high levels in all cells; the levels of a given pol II promoter in a given cell type will depend on the nature of the gene regulatory sequences (enhancers, silencers, *etc.*) present nearby. Prokaryotic RNA polymerase promoters may also be used, providing that the prokaryotic RNA polymerase  
15 enzyme is expressed in the appropriate cells. Ribozymes expressed from such promoters have been shown to function in mammalian cells. Such transcription units can be incorporated into a variety of vectors for introduction into mammalian cells, including but not restricted to, plasmid DNA vectors, viral DNA vectors (such as adenovirus or adeno-associated vectors), or viral RNA vectors (such as retroviral, semliki forest virus, sindbis  
20 virus vectors).

In another embodiment of the invention, peptide nucleic acids (PNAs) compositions are provided. PNA is a DNA mimic in which the nucleobases are attached to a pseudopeptide backbone (Good and Nielsen, *Antisense Nucleic Acid Drug Dev.* 1997 7(4) 431-37). PNA is able to be utilized in a number methods that traditionally have used  
25 RNA or DNA. Often PNA sequences perform better in techniques than the corresponding RNA or DNA sequences and have utilities that are not inherent to RNA or DNA. A review of PNA including methods of making, characteristics of, and methods of using, is provided by Corey (*Trends Biotechnol* 1997 Jun;15(6):224-9). As such, in certain embodiments, one may prepare PNA sequences that are complementary to one or more portions of the ACE

mRNA sequence, and such PNA compositions may be used to regulate, alter, decrease, or reduce the translation of ACE-specific mRNA, and thereby alter the level of ACE activity in a host cell to which such PNA compositions have been administered.

PNAs have 2-aminoethyl-glycine linkages replacing the normal phosphodiester backbone of DNA (Nielsen *et al.*, *Science* 1991 Dec 6;254(5037):1497-500; Hanvey *et al.*, *Science* 1992 Nov 27;258(5087):1481-5; Hyrup and Nielsen, *Bioorg Med Chem.* 1996 Jan;4(1):5-23). This chemistry has three important consequences: firstly, in contrast to DNA or phosphorothioate oligonucleotides, PNAs are neutral molecules; secondly, PNAs are achiral, which avoids the need to develop a stereoselective synthesis; and thirdly, PNA synthesis uses standard Boc or Fmoc protocols for solid-phase peptide synthesis, although other methods, including a modified Merrifield method, have been used.

PNA monomers or ready-made oligomers are commercially available from PerSeptive Biosystems (Framingham, MA). PNA syntheses by either Boc or Fmoc protocols are straightforward using manual or automated protocols (Norton *et al.*, *Bioorg Med Chem.* 1995 Apr;3(4):437-45). The manual protocol lends itself to the production of chemically modified PNAs or the simultaneous synthesis of families of closely related PNAs.

As with peptide synthesis, the success of a particular PNA synthesis will depend on the properties of the chosen sequence. For example, while in theory PNAs can incorporate any combination of nucleotide bases, the presence of adjacent purines can lead to deletions of one or more residues in the product. In expectation of this difficulty, it is suggested that, in producing PNAs with adjacent purines, one should repeat the coupling of residues likely to be added inefficiently. This should be followed by the purification of PNAs by reverse-phase high-pressure liquid chromatography, providing yields and purity of product similar to those observed during the synthesis of peptides.

Modifications of PNAs for a given application may be accomplished by coupling amino acids during solid-phase synthesis or by attaching compounds that contain a carboxylic acid group to the exposed N-terminal amine. Alternatively, PNAs can be

modified after synthesis by coupling to an introduced lysine or cysteine. The ease with which PNAs can be modified facilitates optimization for better solubility or for specific functional requirements. Once synthesized, the identity of PNAs and their derivatives can be confirmed by mass spectrometry. Several studies have made and utilized modifications 5 of PNAs (for example, Norton *et al.*, *Bioorg Med Chem.* 1995 Apr;3(4):437-45; Petersen *et al.*, *J Pept Sci.* 1995 May-Jun;1(3):175-83; Orum *et al.*, *Biotechniques.* 1995 Sep;19(3):472-80; Footer *et al.*, *Biochemistry.* 1996 Aug 20;35(33):10673-9; Griffith *et al.*, *Nucleic Acids Res.* 1995 Aug 11;23(15):3003-8; Pardridge *et al.*, *Proc Natl Acad Sci U S A.* 1995 Jun 6;92(12):5592-6; Boffa *et al.*, *Proc Natl Acad Sci U S A.* 1995 Mar 10 14;92(6):1901-5; Gambacorti-Passerini *et al.*, *Blood.* 1996 Aug 15;88(4):1411-7; Armitage *et al.*, *Proc Natl Acad Sci U S A.* 1997 Nov 11;94(23):12320-5; Seeger *et al.*, *Biotechniques.* 1997 Sep;23(3):512-7). U.S. Patent No. 5,700,922 discusses PNA-DNA-PNA chimeric molecules and their uses in diagnostics, modulating protein in organisms, and treatment of conditions susceptible to therapeutics.

15 Methods of characterizing the antisense binding properties of PNAs are discussed in Rose (*Anal Chem.* 1993 Dec 15;65(24):3545-9) and Jensen *et al.* (*Biochemistry.* 1997 Apr 22;36(16):5072-7). Rose uses capillary gel electrophoresis to determine binding of PNAs to their complementary oligonucleotide, measuring the relative binding kinetics and stoichiometry. Similar types of measurements were made by Jensen *et al.* 20 using BIAcore™ technology.

Other applications of PNAs that have been described and will be apparent to the skilled artisan include use in DNA strand invasion, antisense inhibition, mutational analysis, enhancers of transcription, nucleic acid purification, isolation of transcriptionally active genes, blocking of transcription factor binding, genome cleavage, biosensors, *in situ* 25 hybridization, and the like.

#### Polynucleotide Identification, Characterization and Expression

Polynucleotides compositions of the present invention may be identified, prepared and/or manipulated using any of a variety of well established techniques (see

generally, Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989, and other like references). For example, a polynucleotide may be identified, as described in more detail below, by screening a microarray of cDNAs for tumor-associated expression (*i.e.*, expression that is at least two 5 fold greater in a tumor than in normal tissue, as determined using a representative assay provided herein). Such screens may be performed, for example, using the microarray technology of Affymetrix, Inc. (Santa Clara, CA) according to the manufacturer's instructions (and essentially as described by Schena et al., *Proc. Natl. Acad. Sci. USA* 93:10614-10619, 1996 and Heller et al., *Proc. Natl. Acad. Sci. USA* 94:2150-2155, 1997). 10 Alternatively, polynucleotides may be amplified from cDNA prepared from cells expressing the proteins described herein, such as tumor cells.

Many template dependent processes are available to amplify a target sequences of interest present in a sample. One of the best known amplification methods is the polymerase chain reaction (PCR<sup>TM</sup>) which is described in detail in U.S. Patent Nos. 15 4,683,195, 4,683,202 and 4,800,159, each of which is incorporated herein by reference in its entirety. Briefly, in PCR<sup>TM</sup>, two primer sequences are prepared which are complementary to regions on opposite complementary strands of the target sequence. An excess of deoxynucleoside triphosphates is added to a reaction mixture along with a DNA polymerase (*e.g.*, *Taq* polymerase). If the target sequence is present in a sample, the 20 primers will bind to the target and the polymerase will cause the primers to be extended along the target sequence by adding on nucleotides. By raising and lowering the temperature of the reaction mixture, the extended primers will dissociate from the target to form reaction products, excess primers will bind to the target and to the reaction product and the process is repeated. Preferably reverse transcription and PCR<sup>TM</sup> amplification 25 procedure may be performed in order to quantify the amount of mRNA amplified. Polymerase chain reaction methodologies are well known in the art.

Any of a number of other template dependent processes, many of which are variations of the PCR<sup>TM</sup> amplification technique, are readily known and available in the art. Illustratively, some such methods include the ligase chain reaction (referred to as

LCR), described, for example, in Eur. Pat. Appl. Publ. No. 320,308 and U.S. Patent No. 4,883,750; Qbeta Replicase, described in PCT Intl. Pat. Appl. Publ. No. PCT/US87/00880; Strand Displacement Amplification (SDA) and Repair Chain Reaction (RCR). Still other amplification methods are described in Great Britain Pat. Appl. No. 2 202 328, and in PCT Intl. Pat. Appl. Publ. No. PCT/US89/01025. Other nucleic acid amplification procedures include transcription-based amplification systems (TAS) (PCT Intl. Pat. Appl. Publ. No. WO 88/10315), including nucleic acid sequence based amplification (NASBA) and 3SR. Eur. Pat. Appl. Publ. No. 329,822 describes a nucleic acid amplification process involving cyclically synthesizing single-stranded RNA ("ssRNA"), ssDNA, and double-stranded DNA (dsDNA). PCT Intl. Pat. Appl. Publ. No. WO 89/06700 describes a nucleic acid sequence amplification scheme based on the hybridization of a promoter/primer sequence to a target single-stranded DNA ("ssDNA") followed by transcription of many RNA copies of the sequence. Other amplification methods such as "RACE" (Frohman, 1990), and "one-sided PCR" (Ohara, 1989) are also well-known to those of skill in the art.

An amplified portion of a polynucleotide of the present invention may be used to isolate a full length gene from a suitable library (e.g., a tumor cDNA library) using well known techniques. Within such techniques, a library (cDNA or genomic) is screened using one or more polynucleotide probes or primers suitable for amplification. Preferably, a library is size-selected to include larger molecules. Random primed libraries may also be preferred for identifying 5' and upstream regions of genes. Genomic libraries are preferred for obtaining introns and extending 5' sequences.

For hybridization techniques, a partial sequence may be labeled (e.g., by nick-translation or end-labeling with  $^{32}\text{P}$ ) using well known techniques. A bacterial or bacteriophage library is then generally screened by hybridizing filters containing denatured bacterial colonies (or lawns containing phage plaques) with the labeled probe (see Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989). Hybridizing colonies or plaques are selected and expanded, and the DNA is isolated for further analysis. cDNA clones may be analyzed to determine the amount of additional sequence by, for example, PCR using a primer from

the partial sequence and a primer from the vector. Restriction maps and partial sequences may be generated to identify one or more overlapping clones. The complete sequence may then be determined using standard techniques, which may involve generating a series of deletion clones. The resulting overlapping sequences can then be assembled into a single 5 contiguous sequence. A full length cDNA molecule can be generated by ligating suitable fragments, using well known techniques.

Alternatively, amplification techniques, such as those described above, can be useful for obtaining a full length coding sequence from a partial cDNA sequence. One such amplification technique is inverse PCR (see Triglia et al., *Nucl. Acids Res.* 16:8186, 10 1988), which uses restriction enzymes to generate a fragment in the known region of the gene. The fragment is then circularized by intramolecular ligation and used as a template for PCR with divergent primers derived from the known region. Within an alternative approach, sequences adjacent to a partial sequence may be retrieved by amplification with a primer to a linker sequence and a primer specific to a known region. The amplified 15 sequences are typically subjected to a second round of amplification with the same linker primer and a second primer specific to the known region. A variation on this procedure, which employs two primers that initiate extension in opposite directions from the known sequence, is described in WO 96/38591. Another such technique is known as "rapid amplification of cDNA ends" or RACE. This technique involves the use of an internal 20 primer and an external primer, which hybridizes to a polyA region or vector sequence, to identify sequences that are 5' and 3' of a known sequence. Additional techniques include capture PCR (Lagerstrom et al., *PCR Methods Appl.* 1:111-19, 1991) and walking PCR (Parker et al., *Nucl. Acids. Res.* 19:3055-60, 1991). Other methods employing amplification may also be employed to obtain a full length cDNA sequence.

25 In certain instances, it is possible to obtain a full length cDNA sequence by analysis of sequences provided in an expressed sequence tag (EST) database, such as that available from GenBank. Searches for overlapping ESTs may generally be performed using well known programs (e.g., NCBI BLAST searches), and such ESTs may be used to

generate a contiguous full length sequence. Full length DNA sequences may also be obtained by analysis of genomic fragments.

In other embodiments of the invention, polynucleotide sequences or fragments thereof which encode polypeptides of the invention, or fusion proteins or functional equivalents thereof, may be used in recombinant DNA molecules to direct expression of a polypeptide in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences that encode substantially the same or a functionally equivalent amino acid sequence may be produced and these sequences may be used to clone and express a given polypeptide.

As will be understood by those of skill in the art, it may be advantageous in some instances to produce polypeptide-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce a recombinant RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

Moreover, the polynucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter polypeptide encoding sequences for a variety of reasons, including but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. For example, DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. In addition, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, or introduce mutations, and so forth.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of polypeptide activity, it may be useful to encode a chimeric protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site

located between the polypeptide-encoding sequence and the heterologous protein sequence, so that the polypeptide may be cleaved and purified away from the heterologous moiety.

Sequences encoding a desired polypeptide may be synthesized, in whole or in part, using chemical methods well known in the art (see Caruthers, M. H. et al. (1980) 5 *Nucl. Acids Res. Symp. Ser.* 215-223, Horn, T. et al. (1980) *Nucl. Acids Res. Symp. Ser.* 225-232). Alternatively, the protein itself may be produced using chemical methods to synthesize the amino acid sequence of a polypeptide, or a portion thereof. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge, J. Y. et al. (1995) *Science* 269:202-204) and automated synthesis may be achieved, for example, 10 using the ABI 431A Peptide Synthesizer (Perkin Elmer, Palo Alto, CA).

A newly synthesized peptide may be substantially purified by preparative high performance liquid chromatography (e.g., Creighton, T. (1983) *Proteins, Structures and Molecular Principles*, WH Freeman and Co., New York, N.Y.) or other comparable techniques available in the art. The composition of the synthetic peptides may be 15 confirmed by amino acid analysis or sequencing (e.g., the Edman degradation procedure). Additionally, the amino acid sequence of a polypeptide, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

In order to express a desired polypeptide, the nucleotide sequences encoding 20 the polypeptide, or functional equivalents, may be inserted into appropriate expression vector, *i.e.*, a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence. Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding a polypeptide of interest and appropriate transcriptional and translational control elements. 25 These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. Such techniques are described, for example, in Sambrook, J. et al. (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview, N.Y., and Ausubel, F. M. et al. (1989) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York. N.Y.

A variety of expression vector/host systems may be utilized to contain and express polynucleotide sequences. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems 5 infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems.

The "control elements" or "regulatory sequences" present in an expression 10 vector are those non-translated regions of the vector--enhancers, promoters, 5' and 3' untranslated regions--which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when 15 cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the PBLUESCRIPT phagemid (Stratagene, La Jolla, Calif.) or PSPORT1 plasmid (Gibco BRL, Gaithersburg, MD) and the like may be used. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are generally preferred. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding a polypeptide, 20 vectors based on SV40 or EBV may be advantageously used with an appropriate selectable marker.

In bacterial systems, any of a number of expression vectors may be selected depending upon the use intended for the expressed polypeptide. For example, when large quantities are needed, for example for the induction of antibodies, vectors which direct high 25 level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, the multifunctional *E. coli* cloning and expression vectors such as BLUESCRIPT (Stratagene), in which the sequence encoding the polypeptide of interest may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of .beta.-galactosidase so that a hybrid protein is produced;

pIN vectors (Van Heeke, G. and S. M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509); and the like. pGEX Vectors (Promega, Madison, Wis.) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to 5 glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast, *Saccharomyces cerevisiae*, a number of vectors containing 10 constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used. For reviews, see Ausubel et al. (supra) and Grant et al. (1987) *Methods Enzymol.* 153:516-544.

In cases where plant expression vectors are used, the expression of 15 sequences encoding polypeptides may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311. Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. 20 Cell Differ.* 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (see, for example, Hobbs, S. or Murry, L. E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, N.Y.; pp. 191-196).

25 An insect system may also be used to express a polypeptide of interest. For example, in one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia* larvae. The sequences encoding the polypeptide may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the

polyhedrin promoter. Successful insertion of the polypeptide-encoding sequence will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, *S. frugiperda* cells or *Trichoplusia* larvae in which the polypeptide of interest may be expressed (Engelhard, E. K.

5 et al. (1994) *Proc. Natl. Acad. Sci.* 91:3224-3227).

In mammalian host cells, a number of viral-based expression systems are generally available. For example, in cases where an adenovirus is used as an expression vector, sequences encoding a polypeptide of interest may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing the polypeptide in infected host cells (Logan, J. and Shenk, T. (1984) *Proc. Natl. Acad. Sci.* 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

15 Specific initiation signals may also be used to achieve more efficient translation of sequences encoding a polypeptide of interest. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding the polypeptide, its initiation codon, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a portion thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure 20 translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers which are appropriate for the particular cell system 25 which is used, such as those described in the literature (Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162).

In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired

fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding and/or function. Different host cells such as CHO, COS, HeLa, 5 MDCK, HEK293, and WI38, which have specific cellular machinery and characteristic mechanisms for such post-translational activities, may be chosen to ensure the correct modification and processing of the foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is generally preferred. For example, cell lines which stably express a 10 polynucleotide of interest may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, 15 and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase 20 (Wigler, M. et al. (1977) *Cell* 11:223-32) and adenine phosphoribosyltransferase (Lowy, I. et al. (1990) *Cell* 22:817-23) genes which can be employed in tk.sup.- or aprt.sup.- cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler, M. et al. (1980) *Proc. Natl. Acad. Sci.* 77:3567-70); npt, which confers resistance to the 25 aminoglycosides, neomycin and G-418 (Colbere-Garapin, F. et al (1981) *J. Mol. Biol.* 150:1-14); and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, *supra*). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine (Hartman, S. C. and R. C.

Mulligan (1988) *Proc. Natl. Acad. Sci.* 85:8047-51). The use of visible markers has gained popularity with such markers as anthocyanins, beta-glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression 5 attributable to a specific vector system (Rhodes, C. A. et al. (1995) *Methods Mol. Biol.* 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the sequence encoding a polypeptide is inserted within a marker gene sequence, 10 recombinant cells containing sequences can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a polypeptide-encoding sequence under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells that contain and express a desired polynucleotide 15 sequence may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include, for example, membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein.

20 A variety of protocols for detecting and measuring the expression of polynucleotide-encoded products, using either polyclonal or monoclonal antibodies specific for the product are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to 25 two non-interfering epitopes on a given polypeptide may be preferred for some applications, but a competitive binding assay may also be employed. These and other assays are described, among other places, in Hampton, R. et al. (1990; Serological Methods, a Laboratory Manual, APS Press, St Paul, Minn.) and Maddox, D. E. et al. (1983; *J. Exp. Med.* 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides include oligolabeling, nick translation, end-labeling or PCR amplification 5 using a labeled nucleotide. Alternatively, the sequences, or any portions thereof may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits. Suitable 10 reporter molecules or labels, which may be used include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with a polynucleotide sequence of interest may be cultured under conditions suitable for the expression and recovery of the protein from cell 15 culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides of the invention may be designed to contain signal sequences which direct secretion of the encoded polypeptide through a prokaryotic or eukaryotic cell membrane. Other recombinant constructions may 20 be used to join sequences encoding a polypeptide of interest to nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain 25 utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, Wash.). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, Calif.) between the purification domain and the encoded polypeptide may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing a polypeptide of interest and a

nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilized metal ion affinity chromatography) as described in Porath, J. et al. (1992, *Prot. Exp. Purif.* 3:263-281) while the enterokinase cleavage site provides a means for purifying the desired 5 polypeptide from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D. J. et al. (1993; *DNA Cell Biol.* 12:441-453).

In addition to recombinant production methods, polypeptides of the invention, and fragments thereof, may be produced by direct peptide synthesis using solid-phase techniques (Merrifield J. (1963) *J. Am. Chem. Soc.* 85:2149-2154). Protein synthesis 10 may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Alternatively, various fragments may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

#### Antibody Compositions, Fragments Thereof and Other Binding Agents

15 According to another aspect, the present invention further provides binding agents, such as antibodies and antigen-binding fragments thereof, that exhibit immunological binding to a tumor polypeptide disclosed herein, or to a portion, variant or derivative thereof. An antibody, or antigen-binding fragment thereof, is said to "specifically bind," "immunologically bind," and/or is "immunologically reactive" to a 20 polypeptide of the invention if it reacts at a detectable level (within, for example, an ELISA assay) with the polypeptide, and does not react detectably with unrelated polypeptides under similar conditions.

25 Immunological binding, as used in this context, generally refers to the non-covalent interactions of the type which occur between an immunoglobulin molecule and an antigen for which the immunoglobulin is specific. The strength, or affinity of immunological binding interactions can be expressed in terms of the dissociation constant ( $K_d$ ) of the interaction, wherein a smaller  $K_d$  represents a greater affinity. Immunological binding properties of selected polypeptides can be quantified using methods well known in

the art. One such method entails measuring the rates of antigen-binding site/antigen complex formation and dissociation, wherein those rates depend on the concentrations of the complex partners, the affinity of the interaction, and on geometric parameters that equally influence the rate in both directions. Thus, both the "on rate constant" ( $K_{on}$ ) and the 5 "off rate constant" ( $K_{off}$ ) can be determined by calculation of the concentrations and the actual rates of association and dissociation. The ratio of  $K_{off}/K_{on}$  enables cancellation of all parameters not related to affinity, and is thus equal to the dissociation constant  $K_d$ . See, generally, Davies et al. (1990) *Annual Rev. Biochem.* 59:439-473.

An "antigen-binding site," or "binding portion" of an antibody refers to the 10 part of the immunoglobulin molecule that participates in antigen binding. The antigen binding site is formed by amino acid residues of the N-terminal variable ("V") regions of the heavy ("H") and light ("L") chains. Three highly divergent stretches within the V regions of the heavy and light chains are referred to as "hypervariable regions" which are interposed between more conserved flanking stretches known as "framework regions," or 15 "FRs". Thus the term "FR" refers to amino acid sequences which are naturally found between and adjacent to hypervariable regions in immunoglobulins. In an antibody molecule, the three hypervariable regions of a light chain and the three hypervariable regions of a heavy chain are disposed relative to each other in three dimensional space to form an antigen-binding surface. The antigen-binding surface is complementary to the 20 three-dimensional surface of a bound antigen, and the three hypervariable regions of each of the heavy and light chains are referred to as "complementarity-determining regions," or "CDRs."

Binding agents may be further capable of differentiating between patients with and without a cancer, such as lung cancer, using the representative assays provided 25 herein. For example, antibodies or other binding agents that bind to a tumor protein will preferably generate a signal indicating the presence of a cancer in at least about 20% of patients with the disease, more preferably at least about 30% of patients. Alternatively, or in addition, the antibody will generate a negative signal indicating the absence of the disease in at least about 90% of individuals without the cancer. To determine whether a

binding agent satisfies this requirement, biological samples (e.g., blood, sera, sputum, urine and/or tumor biopsies) from patients with and without a cancer (as determined using standard clinical tests) may be assayed as described herein for the presence of polypeptides that bind to the binding agent. Preferably, a statistically significant number of samples with 5 and without the disease will be assayed. Each binding agent should satisfy the above criteria; however, those of ordinary skill in the art will recognize that binding agents may be used in combination to improve sensitivity.

Any agent that satisfies the above requirements may be a binding agent. For example, a binding agent may be a ribosome, with or without a peptide component, an 10 RNA molecule or a polypeptide. In a preferred embodiment, a binding agent is an antibody or an antigen-binding fragment thereof. Antibodies may be prepared by any of a variety of techniques known to those of ordinary skill in the art. See, e.g., Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In general, antibodies can be produced by cell culture techniques, including the generation of 15 monoclonal antibodies as described herein, or via transfection of antibody genes into suitable bacterial or mammalian cell hosts, in order to allow for the production of recombinant antibodies. In one technique, an immunogen comprising the polypeptide is initially injected into any of a wide variety of mammals (e.g., mice, rats, rabbits, sheep or goats). In this step, the polypeptides of this invention may serve as the immunogen without 20 modification. Alternatively, particularly for relatively short polypeptides, a superior immune response may be elicited if the polypeptide is joined to a carrier protein, such as bovine serum albumin or keyhole limpet hemocyanin. The immunogen is injected into the animal host, preferably according to a predetermined schedule incorporating one or more booster immunizations, and the animals are bled periodically. Polyclonal antibodies 25 specific for the polypeptide may then be purified from such antisera by, for example, affinity chromatography using the polypeptide coupled to a suitable solid support.

Monoclonal antibodies specific for an antigenic polypeptide of interest may be prepared, for example, using the technique of Kohler and Milstein, *Eur. J. Immunol.* 6:511-519, 1976, and improvements thereto. Briefly, these methods involve the

preparation of immortal cell lines capable of producing antibodies having the desired specificity (*i.e.*, reactivity with the polypeptide of interest). Such cell lines may be produced, for example, from spleen cells obtained from an animal immunized as described above. The spleen cells are then immortalized by, for example, fusion with a myeloma cell

5      fusion partner, preferably one that is syngeneic with the immunized animal. A variety of fusion techniques may be employed. For example, the spleen cells and myeloma cells may be combined with a nonionic detergent for a few minutes and then plated at low density on a selective medium that supports the growth of hybrid cells, but not myeloma cells. A preferred selection technique uses HAT (hypoxanthine, aminopterin, thymidine) selection.

10     After a sufficient time, usually about 1 to 2 weeks, colonies of hybrids are observed. Single colonies are selected and their culture supernatants tested for binding activity against the polypeptide. Hybridomas having high reactivity and specificity are preferred.

Monoclonal antibodies may be isolated from the supernatants of growing hybridoma colonies. In addition, various techniques may be employed to enhance the yield, 15 such as injection of the hybridoma cell line into the peritoneal cavity of a suitable vertebrate host, such as a mouse. Monoclonal antibodies may then be harvested from the ascites fluid or the blood. Contaminants may be removed from the antibodies by conventional techniques, such as chromatography, gel filtration, precipitation, and extraction. The polypeptides of this invention may be used in the purification process in, 20 for example, an affinity chromatography step.

A number of therapeutically useful molecules are known in the art which comprise antigen-binding sites that are capable of exhibiting immunological binding properties of an antibody molecule. The proteolytic enzyme papain preferentially cleaves IgG molecules to yield several fragments, two of which (the "F(ab)" fragments) each 25 comprise a covalent heterodimer that includes an intact antigen-binding site. The enzyme pepsin is able to cleave IgG molecules to provide several fragments, including the "F(ab')<sub>2</sub>" fragment which comprises both antigen-binding sites. An "Fv" fragment can be produced by preferential proteolytic cleavage of an IgM, and on rare occasions IgG or IgA immunoglobulin molecule. Fv fragments are, however, more commonly derived using

recombinant techniques known in the art. The Fv fragment includes a non-covalent  $V_H::V_L$  heterodimer including an antigen-binding site which retains much of the antigen recognition and binding capabilities of the native antibody molecule. Inbar et al. (1972) Proc. Nat. Acad. Sci. USA 69:2659-2662; Hochman et al. (1976) Biochem 15:2706-2710; 5 and Ehrlich et al. (1980) Biochem 19:4091-4096.

A single chain Fv ("sFv") polypeptide is a covalently linked  $V_H::V_L$  heterodimer which is expressed from a gene fusion including  $V_H$ - and  $V_L$ -encoding genes linked by a peptide-encoding linker. Huston et al. (1988) Proc. Nat. Acad. Sci. USA 85(16):5879-5883. A number of methods have been described to discern chemical 10 structures for converting the naturally aggregated--but chemically separated--light and heavy polypeptide chains from an antibody V region into an sFv molecule which will fold into a three dimensional structure substantially similar to the structure of an antigen-binding site. See, e.g., U.S. Pat. Nos. 5,091,513 and 5,132,405, to Huston et al.; and U.S. Pat. No. 4,946,778, to Ladner et al.

15 Each of the above-described molecules includes a heavy chain and a light chain CDR set, respectively interposed between a heavy chain and a light chain FR set which provide support to the CDRS and define the spatial relationship of the CDRs relative to each other. As used herein, the term "CDR set" refers to the three hypervariable regions of a heavy or light chain V region. Proceeding from the N-terminus of a heavy or light 20 chain, these regions are denoted as "CDR1," "CDR2," and "CDR3" respectively. An antigen-binding site, therefore, includes six CDRs, comprising the CDR set from each of a heavy and a light chain V region. A polypeptide comprising a single CDR, (e.g., a CDR1, CDR2 or CDR3) is referred to herein as a "molecular recognition unit." Crystallographic analysis of a number of antigen-antibody complexes has demonstrated that the amino acid 25 residues of CDRs form extensive contact with bound antigen, wherein the most extensive antigen contact is with the heavy chain CDR3. Thus, the molecular recognition units are primarily responsible for the specificity of an antigen-binding site.

As used herein, the term "FR set" refers to the four flanking amino acid sequences which frame the CDRs of a CDR set of a heavy or light chain V region. Some

FR residues may contact bound antigen; however, FRs are primarily responsible for folding the V region into the antigen-binding site, particularly the FR residues directly adjacent to the CDRS. Within FRs, certain amino residues and certain structural features are very highly conserved. In this regard, all V region sequences contain an internal disulfide loop of 5 around 90 amino acid residues. When the V regions fold into a binding-site, the CDRs are displayed as projecting loop motifs which form an antigen-binding surface. It is generally recognized that there are conserved structural regions of FRs which influence the folded shape of the CDR loops into certain "canonical" structures--regardless of the precise CDR amino acid sequence. Further, certain FR residues are known to participate in non-covalent 10 interdomain contacts which stabilize the interaction of the antibody heavy and light chains.

A number of "humanized" antibody molecules comprising an antigen-binding site derived from a non-human immunoglobulin have been described, including chimeric antibodies having rodent V regions and their associated CDRs fused to human constant domains (Winter et al. (1991) *Nature* 349:293-299; Lobuglio et al. (1989) *Proc. 15 Nat. Acad. Sci. USA* 86:4220-4224; Shaw et al. (1987) *J Immunol.* 138:4534-4538; and Brown et al. (1987) *Cancer Res.* 47:3577-3583), rodent CDRs grafted into a human supporting FR prior to fusion with an appropriate human antibody constant domain (Riechmann et al. (1988) *Nature* 332:323-327; Verhoeyen et al. (1988) *Science* 239:1534- 20 1536; and Jones et al. (1986) *Nature* 321:522-525), and rodent CDRs supported by recombinantly veneered rodent FRs (European Patent Publication No. 519,596, published Dec. 23, 1992). These "humanized" molecules are designed to minimize unwanted immunological response toward rodent antihuman antibody molecules which limits the duration and effectiveness of therapeutic applications of those moieties in human recipients.

25 As used herein, the terms "veneered FRs" and "recombinantly veneered FRs" refer to the selective replacement of FR residues from, *e.g.*, a rodent heavy or light chain V region, with human FR residues in order to provide a xenogeneic molecule comprising an antigen-binding site which retains substantially all of the native FR polypeptide folding structure. Veneering techniques are based on the understanding that the

ligand binding characteristics of an antigen-binding site are determined primarily by the structure and relative disposition of the heavy and light chain CDR sets within the antigen-binding surface. Davies et al. (1990) Ann. Rev. Biochem. 59:439-473. Thus, antigen binding specificity can be preserved in a humanized antibody only wherein the CDR structures, their interaction with each other, and their interaction with the rest of the V region domains are carefully maintained. By using veneering techniques, exterior (e.g., solvent-accessible) FR residues which are readily encountered by the immune system are selectively replaced with human residues to provide a hybrid molecule that comprises either a weakly immunogenic, or substantially non-immunogenic veneered surface.

The process of veneering makes use of the available sequence data for human antibody variable domains compiled by Kabat et al., in Sequences of Proteins of Immunological Interest, 4th ed., (U.S. Dept. of Health and Human Services, U.S. Government Printing Office, 1987), updates to the Kabat database, and other accessible U.S. and foreign databases (both nucleic acid and protein). Solvent accessibilities of V region amino acids can be deduced from the known three-dimensional structure for human and murine antibody fragments. There are two general steps in veneering a murine antigen-binding site. Initially, the FRs of the variable domains of an antibody molecule of interest are compared with corresponding FR sequences of human variable domains obtained from the above-identified sources. The most homologous human V regions are then compared residue by residue to corresponding murine amino acids. The residues in the murine FR which differ from the human counterpart are replaced by the residues present in the human moiety using recombinant techniques well known in the art. Residue switching is only carried out with moieties which are at least partially exposed (solvent accessible), and care is exercised in the replacement of amino acid residues which may have a significant effect on the tertiary structure of V region domains, such as proline, glycine and charged amino acids.

In this manner, the resultant "veneered" murine antigen-binding sites are thus designed to retain the murine CDR residues, the residues substantially adjacent to the CDRs, the residues identified as buried or mostly buried (solvent inaccessible), the residues

believed to participate in non-covalent (*e.g.*, electrostatic and hydrophobic) contacts between heavy and light chain domains, and the residues from conserved structural regions of the FRs which are believed to influence the "canonical" tertiary structures of the CDR loops. These design criteria are then used to prepare recombinant nucleotide sequences 5 which combine the CDRs of both the heavy and light chain of a murine antigen-binding site into human-appearing FRs that can be used to transfect mammalian cells for the expression of recombinant human antibodies which exhibit the antigen specificity of the murine antibody molecule.

In another embodiment of the invention, monoclonal antibodies of the 10 present invention may be coupled to one or more therapeutic agents. Suitable agents in this regard include radionuclides, differentiation inducers, drugs, toxins, and derivatives thereof. Preferred radionuclides include <sup>90</sup>Y, <sup>123</sup>I, <sup>125</sup>I, <sup>131</sup>I, <sup>186</sup>Re, <sup>188</sup>Re, <sup>211</sup>At, and <sup>212</sup>Bi. Preferred drugs include methotrexate, and pyrimidine and purine analogs. Preferred differentiation inducers include phorbol esters and butyric acid. Preferred toxins include 15 ricin, abrin, diphtheria toxin, cholera toxin, gelonin, *Pseudomonas* exotoxin, *Shigella* toxin, and pokeweed antiviral protein.

A therapeutic agent may be coupled (*e.g.*, covalently bonded) to a suitable monoclonal antibody either directly or indirectly (*e.g.*, via a linker group). A direct reaction 20 between an agent and an antibody is possible when each possesses a substituent capable of reacting with the other. For example, a nucleophilic group, such as an amino or sulphydryl group, on one may be capable of reacting with a carbonyl-containing group, such as an anhydride or an acid halide, or with an alkyl group containing a good leaving group (*e.g.*, a halide) on the other.

Alternatively, it may be desirable to couple a therapeutic agent and an 25 antibody via a linker group. A linker group can function as a spacer to distance an antibody from an agent in order to avoid interference with binding capabilities. A linker group can also serve to increase the chemical reactivity of a substituent on an agent or an antibody, and thus increase the coupling efficiency. An increase in chemical reactivity may also

facilitate the use of agents, or functional groups on agents, which otherwise would not be possible.

It will be evident to those skilled in the art that a variety of bifunctional or polyfunctional reagents, both homo- and hetero-functional (such as those described in the 5 catalog of the Pierce Chemical Co., Rockford, IL), may be employed as the linker group. Coupling may be effected, for example, through amino groups, carboxyl groups, sulphydryl groups or oxidized carbohydrate residues. There are numerous references describing such methodology, *e.g.*, U.S. Patent No. 4,671,958, to Rodwell et al.

Where a therapeutic agent is more potent when free from the antibody 10 portion of the immunoconjugates of the present invention, it may be desirable to use a linker group which is cleavable during or upon internalization into a cell. A number of different cleavable linker groups have been described. The mechanisms for the intracellular release of an agent from these linker groups include cleavage by reduction of a disulfide bond (*e.g.*, U.S. Patent No. 4,489,710, to Spitler), by irradiation of a photolabile 15 bond (*e.g.*, U.S. Patent No. 4,625,014, to Senter et al.), by hydrolysis of derivatized amino acid side chains (*e.g.*, U.S. Patent No. 4,638,045, to Kohn et al.), by serum complement-mediated hydrolysis (*e.g.*, U.S. Patent No. 4,671,958, to Rodwell et al.), and acid-catalyzed hydrolysis (*e.g.*, U.S. Patent No. 4,569,789, to Blattler et al.).

It may be desirable to couple more than one agent to an antibody. In one 20 embodiment, multiple molecules of an agent are coupled to one antibody molecule. In another embodiment, more than one type of agent may be coupled to one antibody. Regardless of the particular embodiment, immunoconjugates with more than one agent may be prepared in a variety of ways. For example, more than one agent may be coupled directly to an antibody molecule, or linkers that provide multiple sites for attachment can be 25 used. Alternatively, a carrier can be used.

A carrier may bear the agents in a variety of ways, including covalent bonding either directly or via a linker group. Suitable carriers include proteins such as albumins (*e.g.*, U.S. Patent No. 4,507,234, to Kato et al.), peptides and polysaccharides such as aminodextran (*e.g.*, U.S. Patent No. 4,699,784, to Shih et al.). A carrier may also

bear an agent by noncovalent bonding or by encapsulation, such as within a liposome vesicle (e.g., U.S. Patent Nos. 4,429,008 and 4,873,088). Carriers specific for radionuclide agents include radiohalogenated small molecules and chelating compounds. For example, U.S. Patent No. 4,735,792 discloses representative radiohalogenated small molecules and 5 their synthesis. A radionuclide chelate may be formed from chelating compounds that include those containing nitrogen and sulfur atoms as the donor atoms for binding the metal, or metal oxide, radionuclide. For example, U.S. Patent No. 4,673,562, to Davison et al. discloses representative chelating compounds and their synthesis.

### T Cell Compositions

10 The present invention, in another aspect, provides T cells specific for a tumor polypeptide disclosed herein, or for a variant or derivative thereof. Such cells may generally be prepared *in vitro* or *ex vivo*, using standard procedures. For example, T cells may be isolated from bone marrow, peripheral blood, or a fraction of bone marrow or peripheral blood of a patient, using a commercially available cell separation system, such as 15 the Isolex<sup>TM</sup> System, available from Nexell Therapeutics, Inc. (Irvine, CA; see also U.S. Patent No. 5,240,856; U.S. Patent No. 5,215,926; WO 89/06280; WO 91/16116 and WO 92/07243). Alternatively, T cells may be derived from related or unrelated humans, non-human mammals, cell lines or cultures.

20 T cells may be stimulated with a polypeptide, polynucleotide encoding a polypeptide and/or an antigen presenting cell (APC) that expresses such a polypeptide. Such stimulation is performed under conditions and for a time sufficient to permit the generation of T cells that are specific for the polypeptide of interest. Preferably, a tumor polypeptide or polynucleotide of the invention is present within a delivery vehicle, such as a microsphere, to facilitate the generation of specific T cells.

25 T cells are considered to be specific for a polypeptide of the present invention if the T cells specifically proliferate, secrete cytokines or kill target cells coated with the polypeptide or expressing a gene encoding the polypeptide. T cell specificity may be evaluated using any of a variety of standard techniques. For example, within a

chromium release assay or proliferation assay, a stimulation index of more than two fold increase in lysis and/or proliferation, compared to negative controls, indicates T cell specificity. Such assays may be performed, for example, as described in Chen et al., *Cancer Res.* 54:1065-1070, 1994. Alternatively, detection of the proliferation of T cells

5 may be accomplished by a variety of known techniques. For example, T cell proliferation can be detected by measuring an increased rate of DNA synthesis (e.g., by pulse-labeling cultures of T cells with tritiated thymidine and measuring the amount of tritiated thymidine incorporated into DNA). Contact with a tumor polypeptide (100 ng/ml - 100  $\mu$ g/ml, preferably 200 ng/ml - 25  $\mu$ g/ml) for 3 - 7 days will typically result in at least a two fold

10 increase in proliferation of the T cells. Contact as described above for 2-3 hours should result in activation of the T cells, as measured using standard cytokine assays in which a two fold increase in the level of cytokine release (e.g., TNF or IFN- $\gamma$ ) is indicative of T cell activation (see Coligan et al., *Current Protocols in Immunology*, vol. 1, Wiley Interscience (Greene 1998)). T cells that have been activated in response to a tumor polypeptide,

15 polynucleotide or polypeptide-expressing APC may be CD4 $^{+}$  and/or CD8 $^{+}$ . Tumor polypeptide-specific T cells may be expanded using standard techniques. Within preferred embodiments, the T cells are derived from a patient, a related donor or an unrelated donor, and are administered to the patient following stimulation and expansion.

For therapeutic purposes, CD4 $^{+}$  or CD8 $^{+}$  T cells that proliferate in response

20 to a tumor polypeptide, polynucleotide or APC can be expanded in number either *in vitro* or *in vivo*. Proliferation of such T cells *in vitro* may be accomplished in a variety of ways. For example, the T cells can be re-exposed to a tumor polypeptide, or a short peptide corresponding to an immunogenic portion of such a polypeptide, with or without the addition of T cell growth factors, such as interleukin-2, and/or stimulator cells that

25 synthesize a tumor polypeptide. Alternatively, one or more T cells that proliferate in the presence of the tumor polypeptide can be expanded in number by cloning. Methods for cloning cells are well known in the art, and include limiting dilution.

### Pharmaceutical Compositions

In additional embodiments, the present invention concerns formulation of one or more of the polynucleotide, polypeptide, T-cell and/or antibody compositions disclosed herein in pharmaceutically-acceptable carriers for administration to a cell or an animal, either alone, or in combination with one or more other modalities of therapy.

It will be understood that, if desired, a composition as disclosed herein may be administered in combination with other agents as well, such as, *e.g.*, other proteins or polypeptides or various pharmaceutically-active agents. In fact, there is virtually no limit to other components that may also be included, given that the additional agents do not cause a significant adverse effect upon contact with the target cells or host tissues. The compositions may thus be delivered along with various other agents as required in the particular instance. Such compositions may be purified from host cells or other biological sources, or alternatively may be chemically synthesized as described herein. Likewise, such compositions may further comprise substituted or derivatized RNA or DNA compositions.

Therefore, in another aspect of the present invention, pharmaceutical compositions are provided comprising one or more of the polynucleotide, polypeptide, antibody, and/or T-cell compositions described herein in combination with a physiologically acceptable carrier. In certain preferred embodiments, the pharmaceutical compositions of the invention comprise immunogenic polynucleotide and/or polypeptide compositions of the invention for use in prophylactic and therapeutic vaccine applications. Vaccine preparation is generally described in, for example, M.F. Powell and M.J. Newman, eds., "Vaccine Design (the subunit and adjuvant approach)," Plenum Press (NY, 1995). Generally, such compositions will comprise one or more polynucleotide and/or polypeptide compositions of the present invention in combination with one or more immunostimulants.

It will be apparent that any of the pharmaceutical compositions described herein can contain pharmaceutically acceptable salts of the polynucleotides and polypeptides of the invention. Such salts can be prepared, for example, from pharmaceutically acceptable non-toxic bases, including organic bases (*e.g.*, salts of primary,

secondary and tertiary amines and basic amino acids) and inorganic bases (e.g., sodium, potassium, lithium, ammonium, calcium and magnesium salts).

In another embodiment, illustrative immunogenic compositions, e.g., vaccine compositions, of the present invention comprise DNA encoding one or more of the 5 polypeptides as described above, such that the polypeptide is generated *in situ*. As noted above, the polynucleotide may be administered within any of a variety of delivery systems known to those of ordinary skill in the art. Indeed, numerous gene delivery techniques are well known in the art, such as those described by Rolland, *Crit. Rev. Therap. Drug Carrier Systems* 15:143-198, 1998, and references cited therein. Appropriate polynucleotide 10 expression systems will, of course, contain the necessary regulatory DNA regulatory sequences for expression in a patient (such as a suitable promoter and terminating signal). Alternatively, bacterial delivery systems may involve the administration of a bacterium (such as *Bacillus-Calmette-Guerrin*) that expresses an immunogenic portion of the polypeptide on its cell surface or secretes such an epitope.

15 Therefore, in certain embodiments, polynucleotides encoding immunogenic polypeptides described herein are introduced into suitable mammalian host cells for expression using any of a number of known viral-based systems. In one illustrative embodiment, retroviruses provide a convenient and effective platform for gene delivery systems. A selected nucleotide sequence encoding a polypeptide of the present invention 20 can be inserted into a vector and packaged in retroviral particles using techniques known in the art. The recombinant virus can then be isolated and delivered to a subject. A number of illustrative retroviral systems have been described (e.g., U.S. Pat. No. 5,219,740; Miller and Rosman (1989) *BioTechniques* 7:980-990; Miller, A. D. (1990) *Human Gene Therapy* 1:5-14; Scarpa et al. (1991) *Virology* 180:849-852; Burns et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:8033-8037; and Boris-Lawrie and Temin (1993) *Cur. Opin. Genet. Develop.* 25 3:102-109.

In addition, a number of illustrative adenovirus-based systems have also been described. Unlike retroviruses which integrate into the host genome, adenoviruses persist extrachromosomally thus minimizing the risks associated with insertional

mutagenesis (Haj-Ahmad and Graham (1986) *J. Virol.* 57:267-274; Bett et al. (1993) *J. Virol.* 67:5911-5921; Mittereder et al. (1994) *Human Gene Therapy* 5:717-729; Seth et al. (1994) *J. Virol.* 68:933-940; Barr et al. (1994) *Gene Therapy* 1:51-58; Berkner, K. L. (1988) *BioTechniques* 6:616-629; and Rich et al. (1993) *Human Gene Therapy* 4:461-476).

5 Various adeno-associated virus (AAV) vector systems have also been developed for polynucleotide delivery. AAV vectors can be readily constructed using techniques well known in the art. See, e.g., U.S. Pat. Nos. 5,173,414 and 5,139,941; International Publication Nos. WO 92/01070 and WO 93/03769; Lebkowski et al. (1988) *Molec. Cell. Biol.* 8:3988-3996; Vincent et al. (1990) *Vaccines* 90 (Cold Spring Harbor 10 Laboratory Press); Carter, B. J. (1992) *Current Opinion in Biotechnology* 3:533-539; Muzyczka, N. (1992) *Current Topics in Microbiol. and Immunol.* 158:97-129; Kotin, R. M. (1994) *Human Gene Therapy* 5:793-801; Shelling and Smith (1994) *Gene Therapy* 1:165-169; and Zhou et al. (1994) *J. Exp. Med.* 179:1867-1875.

15 Additional viral vectors useful for delivering the polynucleotides encoding polypeptides of the present invention by gene transfer include those derived from the pox family of viruses, such as vaccinia virus and avian poxvirus. By way of example, vaccinia virus recombinants expressing the novel molecules can be constructed as follows. The DNA encoding a polypeptide is first inserted into an appropriate vector so that it is adjacent to a vaccinia promoter and flanking vaccinia DNA sequences, such as the sequence 20 encoding thymidine kinase (TK). This vector is then used to transfect cells which are simultaneously infected with vaccinia. Homologous recombination serves to insert the vaccinia promoter plus the gene encoding the polypeptide of interest into the viral genome. The resulting TK.sup.(-) recombinant can be selected by culturing the cells in the presence of 5-bromodeoxyuridine and picking viral plaques resistant thereto.

25 A vaccinia-based infection/transfection system can be conveniently used to provide for inducible, transient expression or coexpression of one or more polypeptides described herein in host cells of an organism. In this particular system, cells are first infected in vitro with a vaccinia virus recombinant that encodes the bacteriophage T7 RNA polymerase. This polymerase displays exquisite specificity in that it only transcribes

templates bearing T7 promoters. Following infection, cells are transfected with the polynucleotide or polynucleotides of interest, driven by a T7 promoter. The polymerase expressed in the cytoplasm from the vaccinia virus recombinant transcribes the transfected DNA into RNA which is then translated into polypeptide by the host translational machinery. The method provides for high level, transient, cytoplasmic production of large quantities of RNA and its translation products. See, e.g., Elroy-Stein and Moss, *Proc. Natl. Acad. Sci. USA* (1990) 87:6743-6747; Fuerst et al. *Proc. Natl. Acad. Sci. USA* (1986) 83:8122-8126.

Alternatively, avipoxviruses, such as the fowlpox and canarypox viruses, can also be used to deliver the coding sequences of interest. Recombinant avipox viruses, expressing immunogens from mammalian pathogens, are known to confer protective immunity when administered to non-avian species. The use of an Avipox vector is particularly desirable in human and other mammalian species since members of the Avipox genus can only productively replicate in susceptible avian species and therefore are not infective in mammalian cells. Methods for producing recombinant Avipoxviruses are known in the art and employ genetic recombination, as described above with respect to the production of vaccinia viruses. See, e.g., WO 91/12882; WO 89/03429; and WO 92/03545.

Any of a number of alphavirus vectors can also be used for delivery of polynucleotide compositions of the present invention, such as those vectors described in U.S. Patent Nos. 5,843,723; 6,015,686; 6,008,035 and 6,015,694. Certain vectors based on Venezuelan Equine Encephalitis (VEE) can also be used, illustrative examples of which can be found in U.S. Patent Nos. 5,505,947 and 5,643,576.

Moreover, molecular conjugate vectors, such as the adenovirus chimeric vectors described in Michael et al. *J. Biol. Chem.* (1993) 268:6866-6869 and Wagner et al. *Proc. Natl. Acad. Sci. USA* (1992) 89:6099-6103, can also be used for gene delivery under the invention.

Additional illustrative information on these and other known viral-based delivery systems can be found, for example, in Fisher-Hoch et al., *Proc. Natl. Acad. Sci. USA* 86:317-321, 1989; Flexner et al., *Ann. N.Y. Acad. Sci.* 569:86-103, 1989; Flexner et

al., *Vaccine* 8:17-21, 1990; U.S. Patent Nos. 4,603,112, 4,769,330, and 5,017,487; WO 89/01973; U.S. Patent No. 4,777,127; GB 2,200,651; EP 0,345,242; WO 91/02805; Berkner, *Biotechniques* 6:616-627, 1988; Rosenfeld et al., *Science* 252:431-434, 1991; Kolls et al., *Proc. Natl. Acad. Sci. USA* 91:215-219, 1994; Kass-Eisler et al., *Proc. Natl. Acad. Sci. USA* 90:11498-11502, 1993; Guzman et al., *Circulation* 88:2838-2848, 1993; and Guzman et al., *Cir. Res.* 73:1202-1207, 1993.

In certain embodiments, a polynucleotide may be integrated into the genome of a target cell. This integration may be in the specific location and orientation *via* homologous recombination (gene replacement) or it may be integrated in a random, non-specific location (gene augmentation). In yet further embodiments, the polynucleotide may be stably maintained in the cell as a separate, episomal segment of DNA. Such polynucleotide segments or "episomes" encode sequences sufficient to permit maintenance and replication independent of or in synchronization with the host cell cycle. The manner in which the expression construct is delivered to a cell and where in the cell the polynucleotide remains is dependent on the type of expression construct employed.

In another embodiment of the invention, a polynucleotide is administered/delivered as "naked" DNA, for example as described in Ulmer et al., *Science* 259:1745-1749, 1993 and reviewed by Cohen, *Science* 259:1691-1692, 1993. The uptake of naked DNA may be increased by coating the DNA onto biodegradable beads, which are efficiently transported into the cells.

In still another embodiment, a composition of the present invention can be delivered via a particle bombardment approach, many of which have been described. In one illustrative example, gas-driven particle acceleration can be achieved with devices such as those manufactured by Powderject Pharmaceuticals PLC (Oxford, UK) and Powderject Vaccines Inc. (Madison, WI), some examples of which are described in U.S. Patent Nos. 5,846,796; 6,010,478; 5,865,796; 5,584,807; and EP Patent No. 0500 799. This approach offers a needle-free delivery approach wherein a dry powder formulation of microscopic particles, such as polynucleotide or polypeptide particles, are accelerated to high speed

within a helium gas jet generated by a hand held device, propelling the particles into a target tissue of interest.

In a related embodiment, other devices and methods that may be useful for gas-driven needle-less injection of compositions of the present invention include those provided by Bioject, Inc. (Portland, OR), some examples of which are described in U.S. Patent Nos. 4,790,824; 5,064,413; 5,312,335; 5,383,851; 5,399,163; 5,520,639 and 5,993,412.

According to another embodiment, the pharmaceutical compositions described herein will comprise one or more immunostimulants in addition to the immunogenic polynucleotide, polypeptide, antibody, T-cell and/or APC compositions of this invention. An immunostimulant refers to essentially any substance that enhances or potentiates an immune response (antibody and/or cell-mediated) to an exogenous antigen. One preferred type of immunostimulant comprises an adjuvant. Many adjuvants contain a substance designed to protect the antigen from rapid catabolism, such as aluminum hydroxide or mineral oil, and a stimulator of immune responses, such as lipid A, *Bordatella pertussis* or *Mycobacterium tuberculosis* derived proteins. Certain adjuvants are commercially available as, for example, Freund's Incomplete Adjuvant and Complete Adjuvant (Difco Laboratories, Detroit, MI); Merck Adjuvant 65 (Merck and Company, Inc., Rahway, NJ); AS-2 (SmithKline Beecham, Philadelphia, PA); aluminum salts such as aluminum hydroxide gel (alum) or aluminum phosphate; salts of calcium, iron or zinc; an insoluble suspension of acylated tyrosine; acylated sugars; cationically or anionically derivatized polysaccharides; polyphosphazenes; biodegradable microspheres; monophosphoryl lipid A and quil A. Cytokines, such as GM-CSF, interleukin-2, -7, -12, and other like growth factors, may also be used as adjuvants.

Within certain embodiments of the invention, the adjuvant composition is preferably one that induces an immune response predominantly of the Th1 type. High levels of Th1-type cytokines (e.g., IFN- $\gamma$ , TNF $\alpha$ , IL-2 and IL-12) tend to favor the induction of cell mediated immune responses to an administered antigen. In contrast, high levels of Th2-type cytokines (e.g., IL-4, IL-5, IL-6 and IL-10) tend to favor the induction of

humoral immune responses. Following application of a vaccine as provided herein, a patient will support an immune response that includes Th1- and Th2-type responses. Within a preferred embodiment, in which a response is predominantly Th1-type, the level of Th1-type cytokines will increase to a greater extent than the level of Th2-type cytokines.

5 The levels of these cytokines may be readily assessed using standard assays. For a review of the families of cytokines, see Mosmann and Coffman, *Ann. Rev. Immunol.* 7:145-173, 1989.

Certain preferred adjuvants for eliciting a predominantly Th1-type response include, for example, a combination of monophosphoryl lipid A, preferably 3-de-O-10 acylated monophosphoryl lipid A, together with an aluminum salt. MPL® adjuvants are available from Corixa Corporation (Seattle, WA; *see*, for example, US Patent Nos. 4,436,727; 4,877,611; 4,866,034 and 4,912,094). CpG-containing oligonucleotides (in which the CpG dinucleotide is unmethylated) also induce a predominantly Th1 response. Such oligonucleotides are well known and are described, for example, in WO 96/02555, 15 WO 99/33488 and U.S. Patent Nos. 6,008,200 and 5,856,462. Immunostimulatory DNA sequences are also described, for example, by Sato et al., *Science* 273:352, 1996. Another preferred adjuvant comprises a saponin, such as Quil A, or derivatives thereof, including QS21 and QS7 (Aquila Biopharmaceuticals Inc., Framingham, MA); Escin; Digitonin; or 20 *Gypsophila* or *Chenopodium quinoa* saponins. Other preferred formulations include more than one saponin in the adjuvant combinations of the present invention, for example combinations of at least two of the following group comprising QS21, QS7, Quil A,  $\beta$ -escin, or digitonin.

Alternatively the saponin formulations may be combined with vaccine vehicles composed of chitosan or other polycationic polymers, polylactide and polylactide-25 co-glycolide particles, poly-N-acetyl glucosamine-based polymer matrix, particles composed of polysaccharides or chemically modified polysaccharides, liposomes and lipid-based particles, particles composed of glycerol monoesters, etc. The saponins may also be formulated in the presence of cholesterol to form particulate structures such as liposomes or ISCOMs. Furthermore, the saponins may be formulated together with a polyoxyethylene

ether or ester, in either a non-particulate solution or suspension, or in a particulate structure such as a paucilamellar liposome or ISCOM. The saponins may also be formulated with excipients such as Carbopol<sup>R</sup> to increase viscosity, or may be formulated in a dry powder form with a powder excipient such as lactose.

5 In one preferred embodiment, the adjuvant system includes the combination of a monophosphoryl lipid A and a saponin derivative, such as the combination of QS21 and 3D-MPL<sup>®</sup> adjuvant, as described in WO 94/00153, or a less reactogenic composition where the QS21 is quenched with cholesterol, as described in WO 96/33739. Other preferred formulations comprise an oil-in-water emulsion and tocopherol. Another 10 particularly preferred adjuvant formulation employing QS21, 3D-MPL<sup>®</sup> adjuvant and tocopherol in an oil-in-water emulsion is described in WO 95/17210.

Another enhanced adjuvant system involves the combination of a CpG-containing oligonucleotide and a saponin derivative particularly the combination of CpG and QS21 is disclosed in WO 00/09159. Preferably the formulation additionally comprises 15 an oil in water emulsion and tocopherol.

Additional illustrative adjuvants for use in the pharmaceutical compositions of the invention include Montanide ISA 720 (Seppic, France), SAF (Chiron, California, United States), ISCOMS (CSL), MF-59 (Chiron), the SBAS series of adjuvants (*e.g.*, SBAS-2 or SBAS-4, available from SmithKline Beecham, Rixensart, Belgium), Detox 20 (Enhanzym<sup>®</sup>) (Corixa, Hamilton, MT), RC-529 (Corixa, Hamilton, MT) and other aminoalkyl glucosaminide 4-phosphates (AGPs), such as those described in pending U.S. Patent Application Serial Nos. 08/853,826 and 09/074,720, the disclosures of which are incorporated herein by reference in their entireties, and polyoxyethylene ether adjuvants such as those described in WO 99/52549A1.

25 Other preferred adjuvants include adjuvant molecules of the general formula (I):  $\text{HO}(\text{CH}_2\text{CH}_2\text{O})_n\text{-A-R}$ ,

wherein,  $n$  is 1-50, A is a bond or  $-\text{C}(\text{O})-$ , R is  $\text{C}_{1-50}$  alkyl or Phenyl  $\text{C}_{1-50}$  alkyl.

One embodiment of the present invention consists of a vaccine formulation comprising a polyoxyethylene ether of general formula (I), wherein  $n$  is between 1 and 50,

preferably 4-24, most preferably 9; the *R* component is C<sub>1-50</sub>, preferably C<sub>4</sub>-C<sub>20</sub> alkyl and most preferably C<sub>12</sub> alkyl, and *A* is a bond. The concentration of the polyoxyethylene ethers should be in the range 0.1-20%, preferably from 0.1-10%, and most preferably in the range 0.1-1%. Preferred polyoxyethylene ethers are selected from the following group: 5 polyoxyethylene-9-lauryl ether, polyoxyethylene-9-stearyl ether, polyoxyethylene-8-stearyl ether, polyoxyethylene-4-lauryl ether, polyoxyethylene-35-lauryl ether, and polyoxyethylene-23-lauryl ether. Polyoxyethylene ethers such as polyoxyethylene lauryl ether are described in the Merck index (12<sup>th</sup> edition: entry 7717). These adjuvant molecules are described in WO 99/52549.

10 The polyoxyethylene ether according to the general formula (I) above may, if desired, be combined with another adjuvant. For example, a preferred adjuvant combination is preferably with CpG as described in the pending UK patent application GB 9820956.2.

According to another embodiment of this invention, an immunogenic 15 composition described herein is delivered to a host via antigen presenting cells (APCs), such as dendritic cells, macrophages, B cells, monocytes and other cells that may be engineered to be efficient APCs. Such cells may, but need not, be genetically modified to increase the capacity for presenting the antigen, to improve activation and/or maintenance of the T cell response, to have anti-tumor effects *per se* and/or to be immunologically 20 compatible with the receiver (*i.e.*, matched HLA haplotype). APCs may generally be isolated from any of a variety of biological fluids and organs, including tumor and peritumoral tissues, and may be autologous, allogeneic, syngeneic or xenogeneic cells.

Certain preferred embodiments of the present invention use dendritic cells or progenitors thereof as antigen-presenting cells. Dendritic cells are highly potent APCs 25 (Banchereau and Steinman, *Nature* 392:245-251, 1998) and have been shown to be effective as a physiological adjuvant for eliciting prophylactic or therapeutic antitumor immunity (see Timmerman and Levy, *Ann. Rev. Med.* 50:507-529, 1999). In general, dendritic cells may be identified based on their typical shape (stellate *in situ*, with marked cytoplasmic processes (dendrites) visible *in vitro*), their ability to take up, process and

present antigens with high efficiency and their ability to activate naïve T cell responses. Dendritic cells may, of course, be engineered to express specific cell-surface receptors or ligands that are not commonly found on dendritic cells *in vivo* or *ex vivo*, and such modified dendritic cells are contemplated by the present invention. As an alternative to 5 dendritic cells, secreted vesicles antigen-loaded dendritic cells (called exosomes) may be used within a vaccine (*see* Zitvogel et al., *Nature Med.* 4:594-600, 1998).

Dendritic cells and progenitors may be obtained from peripheral blood, bone marrow, tumor-infiltrating cells, peritumoral tissues-infiltrating cells, lymph nodes, spleen, skin, umbilical cord blood or any other suitable tissue or fluid. For example, dendritic cells 10 may be differentiated *ex vivo* by adding a combination of cytokines such as GM-CSF, IL-4, IL-13 and/or TNF $\alpha$  to cultures of monocytes harvested from peripheral blood. Alternatively, CD34 positive cells harvested from peripheral blood, umbilical cord blood or 15 bone marrow may be differentiated into dendritic cells by adding to the culture medium combinations of GM-CSF, IL-3, TNF $\alpha$ , CD40 ligand, LPS, flt3 ligand and/or other compound(s) that induce differentiation, maturation and proliferation of dendritic cells.

Dendritic cells are conveniently categorized as "immature" and "mature" cells, which allows a simple way to discriminate between two well characterized phenotypes. However, this nomenclature should not be construed to exclude all possible intermediate stages of differentiation. Immature dendritic cells are characterized as APC 20 with a high capacity for antigen uptake and processing, which correlates with the high expression of Fc $\gamma$  receptor and mannose receptor. The mature phenotype is typically characterized by a lower expression of these markers, but a high expression of cell surface molecules responsible for T cell activation such as class I and class II MHC, adhesion molecules (*e.g.*, CD54 and CD11) and costimulatory molecules (*e.g.*, CD40, CD80, CD86 25 and 4-1BB).

APCs may generally be transfected with a polynucleotide of the invention (or portion or other variant thereof) such that the encoded polypeptide, or an immunogenic portion thereof, is expressed on the cell surface. Such transfection may take place *ex vivo*, and a pharmaceutical composition comprising such transfected cells may then be used for

therapeutic purposes, as described herein. Alternatively, a gene delivery vehicle that targets a dendritic or other antigen presenting cell may be administered to a patient, resulting in transfection that occurs *in vivo*. *In vivo* and *ex vivo* transfection of dendritic cells, for example, may generally be performed using any methods known in the art, such as those 5 described in WO 97/24447, or the gene gun approach described by Mahvi et al., *Immunology and cell Biology* 75:456-460, 1997. Antigen loading of dendritic cells may be achieved by incubating dendritic cells or progenitor cells with the tumor polypeptide, DNA (naked or within a plasmid vector) or RNA; or with antigen-expressing recombinant 10 bacterium or viruses (e.g., vaccinia, fowlpox, adenovirus or lentivirus vectors). Prior to loading, the polypeptide may be covalently conjugated to an immunological partner that provides T cell help (e.g., a carrier molecule). Alternatively, a dendritic cell may be pulsed with a non-conjugated immunological partner, separately or in the presence of the polypeptide.

While any suitable carrier known to those of ordinary skill in the art may be 15 employed in the pharmaceutical compositions of this invention, the type of carrier will typically vary depending on the mode of administration. Compositions of the present invention may be formulated for any appropriate manner of administration, including for example, topical, oral, nasal, mucosal, intravenous, intracranial, intraperitoneal, subcutaneous and intramuscular administration.

20 Carriers for use within such pharmaceutical compositions are biocompatible, and may also be biodegradable. In certain embodiments, the formulation preferably provides a relatively constant level of active component release. In other embodiments, however, a more rapid rate of release immediately upon administration may be desired. The formulation of such compositions is well within the level of ordinary skill in the art 25 using known techniques. Illustrative carriers useful in this regard include microparticles of poly(lactide-co-glycolide), polyacrylate, latex, starch, cellulose, dextran and the like. Other illustrative delayed-release carriers include supramolecular biovectors, which comprise a non-liquid hydrophilic core (e.g., a cross-linked polysaccharide or oligosaccharide) and, optionally, an external layer comprising an amphiphilic compound,

such as a phospholipid (see e.g., U.S. Patent No. 5,151,254 and PCT applications WO 94/20078, WO/94/23701 and WO 96/06638). The amount of active compound contained within a sustained release formulation depends upon the site of implantation, the rate and expected duration of release and the nature of the condition to be treated or prevented.

5 In another illustrative embodiment, biodegradable microspheres (e.g., polylactate polyglycolate) are employed as carriers for the compositions of this invention. Suitable biodegradable microspheres are disclosed, for example, in U.S. Patent Nos. 4,897,268; 5,075,109; 5,928,647; 5,811,128; 5,820,883; 5,853,763; 5,814,344, 5,407,609 and 5,942,252. Modified hepatitis B core protein carrier systems, such as 10 described in WO/99 40934, and references cited therein, will also be useful for many applications. Another illustrative carrier/delivery system employs a carrier comprising particulate-protein complexes, such as those described in U.S. Patent No. 5,928,647, which are capable of inducing a class I-restricted cytotoxic T lymphocyte responses in a host.

15 The pharmaceutical compositions of the invention will often further comprise one or more buffers (e.g., neutral buffered saline or phosphate buffered saline), carbohydrates (e.g., glucose, mannose, sucrose or dextrans), mannitol, proteins, polypeptides or amino acids such as glycine, antioxidants, bacteriostats, chelating agents such as EDTA or glutathione, adjuvants (e.g., aluminum hydroxide), solutes that render the formulation isotonic, hypotonic or weakly hypertonic with the blood of a recipient, 20 suspending agents, thickening agents and/or preservatives. Alternatively, compositions of the present invention may be formulated as a lyophilizate.

25 The pharmaceutical compositions described herein may be presented in unit-dose or multi-dose containers, such as sealed ampoules or vials. Such containers are typically sealed in such a way to preserve the sterility and stability of the formulation until use. In general, formulations may be stored as suspensions, solutions or emulsions in oily or aqueous vehicles. Alternatively, a pharmaceutical composition may be stored in a freeze-dried condition requiring only the addition of a sterile liquid carrier immediately prior to use.

The development of suitable dosing and treatment regimens for using the particular compositions described herein in a variety of treatment regimens, including e.g., oral, parenteral, intravenous, intranasal, and intramuscular administration and formulation, is well known in the art, some of which are briefly discussed below for general purposes of 5. illustration.

In certain applications, the pharmaceutical compositions disclosed herein may be delivered *via* oral administration to an animal. As such, these compositions may be formulated with an inert diluent or with an assimilable edible carrier, or they may be enclosed in hard- or soft-shell gelatin capsule, or they may be compressed into tablets, or 10 they may be incorporated directly with the food of the diet.

The active compounds may even be incorporated with excipients and used in the form of ingestible tablets, buccal tables, troches, capsules, elixirs, suspensions, syrups, wafers, and the like (see, for example, Mathiowitz *et al.*, *Nature* 1997 Mar 27;386(6623):410-4; Hwang *et al.*, *Crit Rev Ther Drug Carrier Syst* 1998;15(3):243-84; U. 15 S. Patent 5,641,515; U. S. Patent 5,580,579 and U. S. Patent 5,792,451). Tablets, troches, pills, capsules and the like may also contain any of a variety of additional components, for example, a binder, such as gum tragacanth, acacia, cornstarch, or gelatin; excipients, such as dicalcium phosphate; a disintegrating agent, such as corn starch, potato starch, alginic acid and the like; a lubricant, such as magnesium stearate; and a sweetening agent, such as 20 sucrose, lactose or saccharin may be added or a flavoring agent, such as peppermint, oil of wintergreen, or cherry flavoring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar, or both. Of course, 25 any material used in preparing any dosage unit form should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compounds may be incorporated into sustained-release preparation and formulations.

Typically, these formulations will contain at least about 0.1% of the active compound or more, although the percentage of the active ingredient(s) may, of course, be

varied and may conveniently be between about 1 or 2% and about 60% or 70% or more of the weight or volume of the total formulation. Naturally, the amount of active compound(s) in each therapeutically useful composition may be prepared in such a way that a suitable dosage will be obtained in any given unit dose of the compound. Factors such as  
5 solubility, bioavailability, biological half-life, route of administration, product shelf life, as well as other pharmacological considerations will be contemplated by one skilled in the art of preparing such pharmaceutical formulations, and as such, a variety of dosages and treatment regimens may be desirable.

For oral administration the compositions of the present invention may  
10 alternatively be incorporated with one or more excipients in the form of a mouthwash, dentifrice, buccal tablet, oral spray, or sublingual orally-administered formulation. Alternatively, the active ingredient may be incorporated into an oral solution such as one containing sodium borate, glycerin and potassium bicarbonate, or dispersed in a dentifrice, or added in a therapeutically-effective amount to a composition that may include water,  
15 binders, abrasives, flavoring agents, foaming agents, and humectants. Alternatively the compositions may be fashioned into a tablet or solution form that may be placed under the tongue or otherwise dissolved in the mouth.

In certain circumstances it will be desirable to deliver the pharmaceutical compositions disclosed herein parenterally, intravenously, intramuscularly, or even  
20 intraperitoneally. Such approaches are well known to the skilled artisan, some of which are further described, for example, in U. S. Patent 5,543,158; U. S. Patent 5,641,515 and U. S. Patent 5,399,363. In certain embodiments, solutions of the active compounds as free base or pharmacologically acceptable salts may be prepared in water suitably mixed with a surfactant, such as hydroxypropylcellulose. Dispersions may also be prepared in glycerol,  
25 liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations generally will contain a preservative to prevent the growth of microorganisms.

Illustrative pharmaceutical forms suitable for injectable use include sterile aqueous solutions or dispersions and sterile powders for the extemporaneous preparation of

sterile injectable solutions or dispersions (for example, see U. S. Patent 5,466,468). In all cases the form must be sterile and must be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms, such as bacteria and fungi. The carrier 5 can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (e.g., glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and/or vegetable oils. Proper fluidity may be maintained, for example, by the use of a coating, such as lecithin, by the maintenance of the required particle size in the case of dispersion and/or by the use of surfactants. The prevention of the action of microorganisms 10 can be facilitated by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and 15 gelatin.

In one embodiment, for parenteral administration in an aqueous solution, the solution should be suitably buffered if necessary and the liquid diluent first rendered isotonic with sufficient saline or glucose. These particular aqueous solutions are especially suitable for intravenous, intramuscular, subcutaneous and intraperitoneal administration. In 20 this connection, a sterile aqueous medium that can be employed will be known to those of skill in the art in light of the present disclosure. For example, one dosage may be dissolved in 1 ml of isotonic NaCl solution and either added to 1000 ml of hypodermoclysis fluid or injected at the proposed site of infusion, (see for example, "Remington's Pharmaceutical Sciences" 15th Edition, pages 1035-1038 and 1570-1580). Some variation in dosage will 25 necessarily occur depending on the condition of the subject being treated. Moreover, for human administration, preparations will of course preferably meet sterility, pyrogenicity, and the general safety and purity standards as required by FDA Office of Biologics standards.

In another embodiment of the invention, the compositions disclosed herein may be formulated in a neutral or salt form. Illustrative pharmaceutically-acceptable salts include the acid addition salts (formed with the free amino groups of the protein) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or 5 such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine and the like. Upon formulation, solutions will be administered in a manner compatible with the dosage formulation and in 10 such amount as is therapeutically effective.

The carriers can further comprise any and all solvents, dispersion media, vehicles, coatings, diluents, antibacterial and antifungal agents, isotonic and absorption delaying agents, buffers, carrier solutions, suspensions, colloids, and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. 15 Except insofar as any conventional media or agent is incompatible with the active ingredient, its use in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions. The phrase "pharmaceutically-acceptable" refers to molecular entities and compositions that do not produce an allergic or similar untoward reaction when administered to a human.

20 In certain embodiments, the pharmaceutical compositions may be delivered by intranasal sprays, inhalation, and/or other aerosol delivery vehicles. Methods for delivering genes, nucleic acids, and peptide compositions directly to the lungs *via* nasal aerosol sprays has been described, *e.g.*, in U. S. Patent 5,756,353 and U. S. Patent 5,804,212. Likewise, the delivery of drugs using intranasal microparticle resins (Takenaga 25 *et al.*, J Controlled Release 1998 Mar 2;52(1-2):81-7) and lysophosphatidyl-glycerol compounds (U. S. Patent 5,725,871) are also well-known in the pharmaceutical arts. Likewise, illustrative transmucosal drug delivery in the form of a polytetrafluoroethylene support matrix is described in U. S. Patent 5,780,045.

In certain embodiments, liposomes, nanocapsules, microparticles, lipid particles, vesicles, and the like, are used for the introduction of the compositions of the present invention into suitable host cells/organisms. In particular, the compositions of the present invention may be formulated for delivery either encapsulated in a lipid particle, a 5 liposome, a vesicle, a nanosphere, or a nanoparticle or the like. Alternatively, compositions of the present invention can be bound, either covalently or non-covalently, to the surface of such carrier vehicles.

The formation and use of liposome and liposome-like preparations as potential drug carriers is generally known to those of skill in the art (see for example, Lasic, 10 Trends Biotechnol 1998 Jul;16(7):307-21; Takakura, Nippon Rinsho 1998 Mar;56(3):691-5; Chandran *et al.*, Indian J Exp Biol. 1997 Aug;35(8):801-9; Margalit, Crit Rev Ther Drug Carrier Syst. 1995;12(2-3):233-61; U.S. Patent 5,567,434; U.S. Patent 5,552,157; U.S. Patent 5,565,213; U.S. Patent 5,738,868 and U.S. Patent 5,795,587, each specifically incorporated herein by reference in its entirety).

15 Liposomes have been used successfully with a number of cell types that are normally difficult to transfect by other procedures, including T cell suspensions, primary hepatocyte cultures and PC 12 cells (Renneisen *et al.*, J Biol Chem. 1990 Sep 25;265(27):16337-42; Muller *et al.*, DNA Cell Biol. 1990 Apr;9(3):221-9). In addition, liposomes are free of the DNA length constraints that are typical of viral-based delivery 20 systems. Liposomes have been used effectively to introduce genes, various drugs, radiotherapeutic agents, enzymes, viruses, transcription factors, allosteric effectors and the like, into a variety of cultured cell lines and animals. Furthermore, the use of liposomes does not appear to be associated with autoimmune responses or unacceptable toxicity after systemic delivery.

25 In certain embodiments, liposomes are formed from phospholipids that are dispersed in an aqueous medium and spontaneously form multilamellar concentric bilayer vesicles (also termed multilamellar vesicles (MLVs)).

Alternatively, in other embodiments, the invention provides for pharmaceutically-acceptable nanocapsule formulations of the compositions of the present

invention. Nanocapsules can generally entrap compounds in a stable and reproducible way (see, for example, Quintanar-Guerrero *et al.*, Drug Dev Ind Pharm. 1998 Dec;24(12):1113-28). To avoid side effects due to intracellular polymeric overloading, such ultrafine particles (sized around 0.1  $\mu\text{m}$ ) may be designed using polymers able to be degraded *in vivo*. Such particles can be made as described, for example, by Couvreur *et al.*, Crit Rev Ther Drug Carrier Syst. 1988;5(1):1-20; zur Muhlen *et al.*, Eur J Pharm Biopharm. 1998 Mar;45(2):149-55; Zambaux *et al.* J Controlled Release. 1998 Jan 2;50(1-3):31-40; and U. S. Patent 5,145,684.

#### Cancer Therapeutic Methods

10 In further aspects of the present invention, the pharmaceutical compositions described herein may be used for the treatment of cancer, particularly for the immunotherapy of lung cancer. Within such methods, the pharmaceutical compositions described herein are administered to a patient, typically a warm-blooded animal, preferably a human. A patient may or may not be afflicted with cancer. Accordingly, the above 15 pharmaceutical compositions may be used to prevent the development of a cancer or to treat a patient afflicted with a cancer. Pharmaceutical compositions and vaccines may be administered either prior to or following surgical removal of primary tumors and/or treatment such as administration of radiotherapy or conventional chemotherapeutic drugs. As discussed above, administration of the pharmaceutical compositions may be by any 20 suitable method, including administration by intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal, intradermal, anal, vaginal, topical and oral routes.

Within certain embodiments, immunotherapy may be active immunotherapy, in which treatment relies on the *in vivo* stimulation of the endogenous host immune system to react against tumors with the administration of immune response-modifying agents (such as polypeptides and polynucleotides as provided herein).

Within other embodiments, immunotherapy may be passive immunotherapy, in which treatment involves the delivery of agents with established tumor-immune reactivity (such as effector cells or antibodies) that can directly or indirectly mediate

antitumor effects and does not necessarily depend on an intact host immune system. Examples of effector cells include T cells as discussed above, T lymphocytes (such as CD8<sup>+</sup> cytotoxic T lymphocytes and CD4<sup>+</sup> T-helper tumor-infiltrating lymphocytes), killer cells (such as Natural Killer cells and lymphokine-activated killer cells), B cells and 5 antigen-presenting cells (such as dendritic cells and macrophages) expressing a polypeptide provided herein. T cell receptors and antibody receptors specific for the polypeptides recited herein may be cloned, expressed and transferred into other vectors or effector cells for adoptive immunotherapy. The polypeptides provided herein may also be used to generate antibodies or anti-idiotypic antibodies (as described above and in U.S. Patent No. 10 4,918,164) for passive immunotherapy.

Effector cells may generally be obtained in sufficient quantities for adoptive immunotherapy by growth *in vitro*, as described herein. Culture conditions for expanding single antigen-specific effector cells to several billion in number with retention of antigen recognition *in vivo* are well known in the art. Such *in vitro* culture conditions typically use 15 intermittent stimulation with antigen, often in the presence of cytokines (such as IL-2) and non-dividing feeder cells. As noted above, immunoreactive polypeptides as provided herein may be used to rapidly expand antigen-specific T cell cultures in order to generate a sufficient number of cells for immunotherapy. In particular, antigen-presenting cells, such as dendritic, macrophage, monocyte, fibroblast and/or B cells, may be pulsed with 20 immunoreactive polypeptides or transfected with one or more polynucleotides using standard techniques well known in the art. For example, antigen-presenting cells can be transfected with a polynucleotide having a promoter appropriate for increasing expression in a recombinant virus or other expression system. Cultured effector cells for use in therapy must be able to grow and distribute widely, and to survive long term *in vivo*. 25 Studies have shown that cultured effector cells can be induced to grow *in vivo* and to survive long term in substantial numbers by repeated stimulation with antigen supplemented with IL-2 (see, for example, Cheever et al., *Immunological Reviews* 157:177, 1997).

Alternatively, a vector expressing a polypeptide recited herein may be introduced into antigen presenting cells taken from a patient and clonally propagated *ex vivo* for transplant back into the same patient. Transfected cells may be reintroduced into the patient using any means known in the art, preferably in sterile form by intravenous, 5 intracavitory, intraperitoneal or intratumor administration.

Routes and frequency of administration of the therapeutic compositions described herein, as well as dosage, will vary from individual to individual, and may be readily established using standard techniques. In general, the pharmaceutical compositions and vaccines may be administered by injection (e.g., intracutaneous, intramuscular, 10 intravenous or subcutaneous), intranasally (e.g., by aspiration) or orally. Preferably, between 1 and 10 doses may be administered over a 52 week period. Preferably, 6 doses are administered, at intervals of 1 month, and booster vaccinations may be given periodically thereafter. Alternate protocols may be appropriate for individual patients. A suitable dose is an amount of a compound that, when administered as described above, is 15 capable of promoting an anti-tumor immune response, and is at least 10-50% above the basal (i.e., untreated) level. Such response can be monitored by measuring the anti-tumor antibodies in a patient or by vaccine-dependent generation of cytolytic effector cells capable of killing the patient's tumor cells *in vitro*. Such vaccines should also be capable of causing an immune response that leads to an improved clinical outcome (e.g., more 20 frequent remissions, complete or partial or longer disease-free survival) in vaccinated patients as compared to non-vaccinated patients. In general, for pharmaceutical compositions and vaccines comprising one or more polypeptides, the amount of each polypeptide present in a dose ranges from about 25 µg to 5 mg per kg of host. Suitable dose sizes will vary with the size of the patient, but will typically range from about 0.1 mL 25 to about 5 mL.

In general, an appropriate dosage and treatment regimen provides the active compound(s) in an amount sufficient to provide therapeutic and/or prophylactic benefit. Such a response can be monitored by establishing an improved clinical outcome (e.g., more frequent remissions, complete or partial, or longer disease-free survival) in treated patients

as compared to non-treated patients. Increases in preexisting immune responses to a tumor protein generally correlate with an improved clinical outcome. Such immune responses may generally be evaluated using standard proliferation, cytotoxicity or cytokine assays, which may be performed using samples obtained from a patient before and after treatment.

5 Cancer Detection and Diagnostic Compositions, Methods and Kits

In general, a cancer may be detected in a patient based on the presence of one or more lung tumor proteins and/or polynucleotides encoding such proteins in a biological sample (for example, blood, sera, sputum urine and/or tumor biopsies) obtained from the patient. In other words, such proteins may be used as markers to indicate the 10 presence or absence of a cancer such as lung cancer. In addition, such proteins may be useful for the detection of other cancers. The binding agents provided herein generally permit detection of the level of antigen that binds to the agent in the biological sample. Polynucleotide primers and probes may be used to detect the level of mRNA encoding a tumor protein, which is also indicative of the presence or absence of a cancer. In general, a 15 lung tumor sequence should be present at a level that is at least three fold higher in tumor tissue than in normal tissue

There are a variety of assay formats known to those of ordinary skill in the art for using a binding agent to detect polypeptide markers in a sample. *See, e.g., Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, 1988.* In 20 general, the presence or absence of a cancer in a patient may be determined by (a) contacting a biological sample obtained from a patient with a binding agent; (b) detecting in the sample a level of polypeptide that binds to the binding agent; and (c) comparing the level of polypeptide with a predetermined cut-off value.

In a preferred embodiment, the assay involves the use of binding agent 25 immobilized on a solid support to bind to and remove the polypeptide from the remainder of the sample. The bound polypeptide may then be detected using a detection reagent that contains a reporter group and specifically binds to the binding agent/polypeptide complex. Such detection reagents may comprise, for example, a binding agent that specifically binds

to the polypeptide or an antibody or other agent that specifically binds to the binding agent, such as an anti-immunoglobulin, protein G, protein A or a lectin. Alternatively, a competitive assay may be utilized, in which a polypeptide is labeled with a reporter group and allowed to bind to the immobilized binding agent after incubation of the binding agent with the sample. The extent to which components of the sample inhibit the binding of the labeled polypeptide to the binding agent is indicative of the reactivity of the sample with the immobilized binding agent. Suitable polypeptides for use within such assays include full length lung tumor proteins and polypeptide portions thereof to which the binding agent binds, as described above.

The solid support may be any material known to those of ordinary skill in the art to which the tumor protein may be attached. For example, the solid support may be a test well in a microtiter plate or a nitrocellulose or other suitable membrane. Alternatively, the support may be a bead or disc, such as glass, fiberglass, latex or a plastic material such as polystyrene or polyvinylchloride. The support may also be a magnetic particle or a fiber optic sensor, such as those disclosed, for example, in U.S. Patent No. 5,359,681. The binding agent may be immobilized on the solid support using a variety of techniques known to those of skill in the art, which are amply described in the patent and scientific literature. In the context of the present invention, the term "immobilization" refers to both noncovalent association, such as adsorption, and covalent attachment (which may be a direct linkage between the agent and functional groups on the support or may be a linkage by way of a cross-linking agent). Immobilization by adsorption to a well in a microtiter plate or to a membrane is preferred. In such cases, adsorption may be achieved by contacting the binding agent, in a suitable buffer, with the solid support for a suitable amount of time. The contact time varies with temperature, but is typically between about 1 hour and about 1 day. In general, contacting a well of a plastic microtiter plate (such as polystyrene or polyvinylchloride) with an amount of binding agent ranging from about 10 ng to about 10  $\mu$ g, and preferably about 100 ng to about 1  $\mu$ g, is sufficient to immobilize an adequate amount of binding agent.

Covalent attachment of binding agent to a solid support may generally be achieved by first reacting the support with a bifunctional reagent that will react with both the support and a functional group, such as a hydroxyl or amino group, on the binding agent. For example, the binding agent may be covalently attached to supports having an appropriate polymer coating using benzoquinone or by condensation of an aldehyde group on the support with an amine and an active hydrogen on the binding partner (see, e.g., Pierce Immunotechnology Catalog and Handbook, 1991, at A12-A13).

In certain embodiments, the assay is a two-antibody sandwich assay. This assay may be performed by first contacting an antibody that has been immobilized on a solid support, commonly the well of a microtiter plate, with the sample, such that polypeptides within the sample are allowed to bind to the immobilized antibody. Unbound sample is then removed from the immobilized polypeptide-antibody complexes and a detection reagent (preferably a second antibody capable of binding to a different site on the polypeptide) containing a reporter group is added. The amount of detection reagent that remains bound to the solid support is then determined using a method appropriate for the specific reporter group.

More specifically, once the antibody is immobilized on the support as described above, the remaining protein binding sites on the support are typically blocked. Any suitable blocking agent known to those of ordinary skill in the art, such as bovine serum albumin or Tween 20<sup>TM</sup> (Sigma Chemical Co., St. Louis, MO). The immobilized antibody is then incubated with the sample, and polypeptide is allowed to bind to the antibody. The sample may be diluted with a suitable diluent, such as phosphate-buffered saline (PBS) prior to incubation. In general, an appropriate contact time (*i.e.*, incubation time) is a period of time that is sufficient to detect the presence of polypeptide within a sample obtained from an individual with lung cancer. Preferably, the contact time is sufficient to achieve a level of binding that is at least about 95% of that achieved at equilibrium between bound and unbound polypeptide. Those of ordinary skill in the art will recognize that the time necessary to achieve equilibrium may be readily determined by

assaying the level of binding that occurs over a period of time. At room temperature, an incubation time of about 30 minutes is generally sufficient.

Unbound sample may then be removed by washing the solid support with an appropriate buffer, such as PBS containing 0.1% Tween 20<sup>TM</sup>. The second antibody, which 5 contains a reporter group, may then be added to the solid support. Preferred reporter groups include those groups recited above.

The detection reagent is then incubated with the immobilized antibody-polypeptide complex for an amount of time sufficient to detect the bound polypeptide. An appropriate amount of time may generally be determined by assaying the level of binding 10 that occurs over a period of time. Unbound detection reagent is then removed and bound detection reagent is detected using the reporter group. The method employed for detecting the reporter group depends upon the nature of the reporter group. For radioactive groups, scintillation counting or autoradiographic methods are generally appropriate. Spectroscopic methods may be used to detect dyes, luminescent groups and fluorescent groups. Biotin 15 may be detected using avidin, coupled to a different reporter group (commonly a radioactive or fluorescent group or an enzyme). Enzyme reporter groups may generally be detected by the addition of substrate (generally for a specific period of time), followed by spectroscopic or other analysis of the reaction products.

To determine the presence or absence of a cancer, such as lung cancer, the 20 signal detected from the reporter group that remains bound to the solid support is generally compared to a signal that corresponds to a predetermined cut-off value. In one preferred embodiment, the cut-off value for the detection of a cancer is the average mean signal obtained when the immobilized antibody is incubated with samples from patients without the cancer. In general, a sample generating a signal that is three standard deviations above 25 the predetermined cut-off value is considered positive for the cancer. In an alternate preferred embodiment, the cut-off value is determined using a Receiver Operator Curve, according to the method of Sackett et al., *Clinical Epidemiology: A Basic Science for Clinical Medicine*, Little Brown and Co., 1985, p. 106-7. Briefly, in this embodiment, the cut-off value may be determined from a plot of pairs of true positive rates (i.e., sensitivity)

and false positive rates (100%-specificity) that correspond to each possible cut-off value for the diagnostic test result. The cut-off value on the plot that is the closest to the upper left-hand corner (*i.e.*, the value that encloses the largest area) is the most accurate cut-off value, and a sample generating a signal that is higher than the cut-off value determined by this 5 method may be considered positive. Alternatively, the cut-off value may be shifted to the left along the plot, to minimize the false positive rate, or to the right, to minimize the false negative rate. In general, a sample generating a signal that is higher than the cut-off value determined by this method is considered positive for a cancer.

In a related embodiment, the assay is performed in a flow-through or strip 10 test format, wherein the binding agent is immobilized on a membrane, such as nitrocellulose. In the flow-through test, polypeptides within the sample bind to the immobilized binding agent as the sample passes through the membrane. A second, labeled binding agent then binds to the binding agent-polypeptide complex as a solution containing the second binding agent flows through the membrane. The detection of bound second 15 binding agent may then be performed as described above. In the strip test format, one end of the membrane to which binding agent is bound is immersed in a solution containing the sample. The sample migrates along the membrane through a region containing second binding agent and to the area of immobilized binding agent. Concentration of second binding agent at the area of immobilized antibody indicates the presence of a cancer. 20 Typically, the concentration of second binding agent at that site generates a pattern, such as a line, that can be read visually. The absence of such a pattern indicates a negative result. In general, the amount of binding agent immobilized on the membrane is selected to generate a visually discernible pattern when the biological sample contains a level of polypeptide that would be sufficient to generate a positive signal in the two-antibody 25 sandwich assay, in the format discussed above. Preferred binding agents for use in such assays are antibodies and antigen-binding fragments thereof. Preferably, the amount of antibody immobilized on the membrane ranges from about 25 ng to about 1 $\mu$ g, and more preferably from about 50 ng to about 500 ng. Such tests can typically be performed with a very small amount of biological sample.

Of course, numerous other assay protocols exist that are suitable for use with the tumor proteins or binding agents of the present invention. The above descriptions are intended to be exemplary only. For example, it will be apparent to those of ordinary skill in the art that the above protocols may be readily modified to use tumor polypeptides 5 to detect antibodies that bind to such polypeptides in a biological sample. The detection of such tumor protein specific antibodies may correlate with the presence of a cancer.

A cancer may also, or alternatively, be detected based on the presence of T cells that specifically react with a tumor protein in a biological sample. Within certain methods, a biological sample comprising CD4<sup>+</sup> and/or CD8<sup>+</sup> T cells isolated from a patient 10 is incubated with a tumor polypeptide, a polynucleotide encoding such a polypeptide and/or an APC that expresses at least an immunogenic portion of such a polypeptide, and the presence or absence of specific activation of the T cells is detected. Suitable biological samples include, but are not limited to, isolated T cells. For example, T cells may be isolated from a patient by routine techniques (such as by Ficoll/Hypaque density gradient 15 centrifugation of peripheral blood lymphocytes). T cells may be incubated *in vitro* for 2-9 days (typically 4 days) at 37°C with polypeptide (e.g., 5 - 25 µg/ml). It may be desirable to incubate another aliquot of a T cell sample in the absence of tumor polypeptide to serve as a control. For CD4<sup>+</sup> T cells, activation is preferably detected by evaluating proliferation of 20 the T cells. For CD8<sup>+</sup> T cells, activation is preferably detected by evaluating cytolytic activity. A level of proliferation that is at least two fold greater and/or a level of cytolytic activity that is at least 20% greater than in disease-free patients indicates the presence of a cancer in the patient.

As noted above, a cancer may also, or alternatively, be detected based on the level of mRNA encoding a tumor protein in a biological sample. For example, at least two 25 oligonucleotide primers may be employed in a polymerase chain reaction (PCR) based assay to amplify a portion of a tumor cDNA derived from a biological sample, wherein at least one of the oligonucleotide primers is specific for (*i.e.*, hybridizes to) a polynucleotide encoding the tumor protein. The amplified cDNA is then separated and detected using techniques well known in the art, such as gel electrophoresis. Similarly, oligonucleotide

probes that specifically hybridize to a polynucleotide encoding a tumor protein may be used in a hybridization assay to detect the presence of polynucleotide encoding the tumor protein in a biological sample.

To permit hybridization under assay conditions, oligonucleotide primers and probes should comprise an oligonucleotide sequence that has at least about 60%, preferably at least about 75% and more preferably at least about 90%, identity to a portion of a polynucleotide encoding a tumor protein of the invention that is at least 10 nucleotides, and preferably at least 20 nucleotides, in length. Preferably, oligonucleotide primers and/or probes hybridize to a polynucleotide encoding a polypeptide described herein under moderately stringent conditions, as defined above. Oligonucleotide primers and/or probes which may be usefully employed in the diagnostic methods described herein preferably are at least 10-40 nucleotides in length. In a preferred embodiment, the oligonucleotide primers comprise at least 10 contiguous nucleotides, more preferably at least 15 contiguous nucleotides, of a DNA molecule having a sequence as disclosed herein. Techniques for both PCR based assays and hybridization assays are well known in the art (see, for example, Mullis et al., *Cold Spring Harbor Symp. Quant. Biol.*, 51:263, 1987; Erlich ed., *PCR Technology*, Stockton Press, NY, 1989).

One preferred assay employs RT-PCR, in which PCR is applied in conjunction with reverse transcription. Typically, RNA is extracted from a biological sample, such as biopsy tissue, and is reverse transcribed to produce cDNA molecules. PCR amplification using at least one specific primer generates a cDNA molecule, which may be separated and visualized using, for example, gel electrophoresis. Amplification may be performed on biological samples taken from a test patient and from an individual who is not afflicted with a cancer. The amplification reaction may be performed on several dilutions of cDNA spanning two orders of magnitude. A two-fold or greater increase in expression in several dilutions of the test patient sample as compared to the same dilutions of the non-cancerous sample is typically considered positive.

In another embodiment, the compositions described herein may be used as markers for the progression of cancer. In this embodiment, assays as described above for

the diagnosis of a cancer may be performed over time, and the change in the level of reactive polypeptide(s) or polynucleotide(s) evaluated. For example, the assays may be performed every 24-72 hours for a period of 6 months to 1 year, and thereafter performed as needed. In general, a cancer is progressing in those patients in whom the level of 5 polypeptide or polynucleotide detected increases over time. In contrast, the cancer is not progressing when the level of reactive polypeptide or polynucleotide either remains constant or decreases with time.

Certain *in vivo* diagnostic assays may be performed directly on a tumor. One such assay involves contacting tumor cells with a binding agent. The bound binding 10 agent may then be detected directly or indirectly via a reporter group. Such binding agents may also be used in histological applications. Alternatively, polynucleotide probes may be used within such applications.

As noted above, to improve sensitivity, multiple tumor protein markers may be assayed within a given sample. It will be apparent that binding agents specific for 15 different proteins provided herein may be combined within a single assay. Further, multiple primers or probes may be used concurrently. The selection of tumor protein markers may be based on routine experiments to determine combinations that results in optimal sensitivity. In addition, or alternatively, assays for tumor proteins provided herein may be combined with assays for other known tumor antigens.

20 The present invention further provides kits for use within any of the above diagnostic methods. Such kits typically comprise two or more components necessary for performing a diagnostic assay. Components may be compounds, reagents, containers and/or equipment. For example, one container within a kit may contain a monoclonal antibody or fragment thereof that specifically binds to a tumor protein. Such antibodies or 25 fragments may be provided attached to a support material, as described above. One or more additional containers may enclose elements, such as reagents or buffers, to be used in the assay. Such kits may also, or alternatively, contain a detection reagent as described above that contains a reporter group suitable for direct or indirect detection of antibody binding.

Alternatively, a kit may be designed to detect the level of mRNA encoding a tumor protein in a biological sample. Such kits generally comprise at least one oligonucleotide probe or primer, as described above, that hybridizes to a polynucleotide encoding a tumor protein. Such an oligonucleotide may be used, for example, within a 5 PCR or hybridization assay. Additional components that may be present within such kits include a second oligonucleotide and/or a diagnostic reagent or container to facilitate the detection of a polynucleotide encoding a tumor protein.

The following Examples are offered by way of illustration and not by way of limitation.

10

#### EXAMPLE 1

#### USE OF MOUSE ANTISERA TO IDENTIFY cDNA SEQUENCES ENCODING LUNG TUMOR ANTIGENS

This example illustrates the isolation of cDNA sequences encoding lung 15 tumor antigens by screening of lung tumor cDNA libraries with mouse anti-tumor sera.

A small cell cDNA lung tumor expression library was constructed using mRNA from the small cell carcinoma cell line NCIH69 employing the Lambda ZAP Express expression system (Stratagene, La Jolla, CA). Mouse anti-SCID mouse serum was developed by growing the lung small cell carcinoma cell lines NCIH69 and NCIH128 in 20 SCID mice, removed SCID serum containing shed and secreted tumor antigens. These sera were pooled and injected into normal mice to produce anti-lung carcinoma serum. The antiserum was adsorbed with *E. coli* lysate and human GAPDH protein, and human PBMC lysate was added to the serum to block antibody to proteins found in normal tissue. The cDNA expression library was then screened with this anti-serum using a goat anti-mouse 25 IgG-A-M (H+L) alkaline phosphatase second antibody developed with NBT/BCIP (Gibco BRL Labs., Gaithersburg, MD). Phage was purified and phagemid excised for clones with inserts in a pBK-CMV vector for expression in prokaryotic or eukaryotic cells.

The determined cDNA sequences for 76 isolated clones are provided in SEQ ID NO:1-76. Comparison of these sequences with those in the public database as described

above, revealed no significant homologies to SEQ ID NO:7, 14, 21, 46 and 55. SEQ ID NO:11, 16, 20, 41, 49 and 74 were found to show some homology to previously identified expressed sequence tags (ESTs). The remaining clones were found to show some degree of homology to previously identified genes. The expression levels of certain of the isolated 5 antigens in lung tumor tissues compared to expression levels in 36 normal tissues was determined by microarray technology. The results of these studies are shown below in Table 2, together with the database analyses for these sequences.

**Table 2**

10

			Lung Tumor Over-Expression ( $\geq 2$ )			
Clone Name	SEQ ID NO:	Description	LT+F/N	SCC+M/N	Squa/N	Aden/N
LSC-2	2	CDM, 6C6, BAP31/2	-	-	2.4	-
LSC-6	5	Motor protein p87/89	-	2.2	-	-
LSC-7	6	Ku autoantien 70 kDa	-	2.2	2.3	-
LSC-10	8	PIBF1 protein	-	2.4	-	-
LSC-11	9	Ku autoantien 70 kDa	-	2.4	-	-
LSC-15	11	Novel	-	-	2.9	-
LSC-29	19	Unknown DKFZp586N1020	-	2.5	-	-
LSC-33	22	10 methylene 4 hydrofolate DH	-	2.4	-	-
LSC-39	26	P1 protein	2.4	5.0	2.8	-
LSC-43	27	Minichrom maint deficient	2.3	7.1	2.8	-
LSC-46	28	Non-metastatic cell 1 NME1	2.6	2.5	2.7	2.1
LSC-49	29	GTPase act. Pro. ID- GAP	3.6	10.0	3.8	3.2
LSC-51	30	ZIC family member 2	-	3.4	-	-
LSC-55	32	Transmembrane(63 kDa) ER	2.7	2.2	4.2	2.3
LSC-64	35	Macrophage Migr Inhib Fac	2.6	3.2	3.9	-
LSC-72	38	hRif beta (p102 protein)	2.4	7.0	2.6	-
LSC-76	40	Pro Synth Init. Factor	-	-	-	2.1

LSC-78	42	Motor protein p87/89	-	2.7	2.2	-
LSC-81	43	Epidermal GFR subst 8 EPS8	-	-	2.7	2.1
LSC-101	45	Transmembrane(63 kDa) ER	2.7	-	4.3	-
LSC-103	47	Nuclear factor 4	-	4.3	2.8	-
LSC-134	51	Fumarase	-	3.6	-	-
LSC-142	52	Unknown BAC CTA363M4	-	-	2.5	-
LSC-144	53	Accessory Pro BAP31/BAP2	2.5	-	2.9	2.4
LSC-148	54	Unknown DKFZp586N1020	-	2.2	-	-
LSC-149	55	Novel / Novel	2.6	2.4	3.1	3.5
LSC-163	57	Unknown Ch8p11.2 sect2/19	-	-	-	2.4
LSC-170	58	Unknown PAC DJ0777023	2.2	3.0	2.2	-
LSC-210	74	Novel	-	2.6	2.1	-

LT+F/N = Lung Tumor plus Fetal tissue over Normal tissues

SC+M/N = Lung Small Cell carcinoma plus Metastatic over Normal tissues

Squa/N = Squamous lung tumor over Normal tissues

5 Aden/N = Adenocarcinoma over Normal tissues

## EXAMPLE 2

### ISOLATION OF LUNG TUMOR cDNA SEQUENCES BY CONVENTIONAL SUBTRACTION

10 A human lung squamous cell carcinoma cDNA expression library was constructed from poly A<sup>+</sup> RNA from a pool of two patient tissues using a Superscript Plasmid System for cDNA Synthesis and Plasmid Cloning kit (BRL Life Technologies, Gaithersburg, MD) following the manufacturer's protocol. Specifically, lung carcinoma tissues were homogenized with polytron (Kinematica, Switzerland) and total RNA was 15 extracted using Trizol reagent (BRL Life Technologies) as directed by the manufacturer. The poly A<sup>+</sup> RNA was then purified using an oligo dT cellulose column as described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989. First-strand cDNA was synthesized using the

NotI/Oligo-dT18 primer. Double-stranded cDNA was synthesized, ligated with BstXI/EcoRI adaptors (Invitrogen, San Diego, CA) and digested with NotI. Following size fractionation with cDNA size fractionation columns (BRL Life Technologies), the cDNA was ligated into the BstXI/NotI site of pcDNA3.1 (Invitrogen) and transformed into 5 ElectroMax *E. coli* DH10B cells (BRL Life Technologies) by electroporation.

Using the same procedure, a normal human lung cDNA expression library was prepared from a pool of normal lung, kidney, colon, pancreas, brain, resting PBMC, heart, skin and esophagus, with esophagus cDNAs making up one third of the material. For both libraries, sequence analysis showed that the majority of clones had a full length cDNA 10 sequence and were synthesized from mRNA

cDNA library subtraction was performed using the above lung squamous cell carcinoma and normal cDNA library, as described by Hara *et al.* (*Blood*, 84:189-199, 1994) with some modifications. Specifically, a lung squamous cell carcinoma-specific subtracted cDNA library was generated as follows. To from the driver cDNA, normal 15 tissue cDNA library (80 µg) was digested with BamHI and XhoI, followed by a filling-in reaction with DNA polymerase Klenow fragment. After phenol-chloroform extraction and ethanol precipitation, the DNA was dissolved in 133 µl of H<sub>2</sub>O, heat-denatured and mixed with 133 µl (133 µg) of Photoprobe biotin (Vector Laboratories, Burlingame, CA). As recommended by the manufacturer, the resulting mixture was irradiated with a 270 W 20 sunlamp on ice for 20 minutes. Additional Photoprobe biotin (67 µl) was added and the biotinylation reaction was repeated. After extraction with butanol five times, the DNA was ethanol-precipitated and dissolved in 23 µl H<sub>2</sub>O to form the driver DNA.

To form the tracer DNA, 10 µg lung squamous cell carcinoma cDNA library was digested with NotI and SpeI, phenol chloroform extracted and passed through Chroma 25 spin-400 columns (Clontech, Palo Alto, CA). Typically, 5 µg of cDNA was recovered after the sizing column. Following ethanol precipitation, the tracer DNA was dissolved in 5 µl H<sub>2</sub>O. Tracer DNA was mixed with 15 µl driver DNA and 20 µl of 2 x hybridization buffer (1.5 M NaCl/10 mM EDTA/50 mM HEPES pH 7.5/0.2% sodium dodecyl sulfate), overlaid with mineral oil, and heat-denatured completely. The sample was immediately transferred

into a 68 °C water bath and incubated for 20 hours (long hybridization [LH]). The reaction mixture was then subjected to a streptavidin treatment followed by phenol/chloroform extraction. This process was repeated three more times. Subtracted DNA was precipitated, dissolved in 12 µl H<sub>2</sub>O, mixed with 8 µl driver DNA and 20 µl of 2 x hybridization buffer, 5 and subjected to a hybridization at 68 °C for 2 hours (short hybridization [SH]). After removal of biotinylated double-stranded DNA, subtracted cDNA was ligated into NotI/SpeI site of chloramphenicol resistant pBCSK<sup>+</sup> (Stratagene, La Jolla, CA) and transformed into ElectroMax *E. coli* DH10B cells by electroporation to generate a lung squamous cell carcinoma specific subtracted cDNA library (referred to as LST-69).

10 A cDNA library (referred to as mets3616A) was constructed from a metastatic lung adenocarcinoma. The mets3616A cDNA library was subtracted against a cDNA library prepared from a pool of normal lung, liver, pancreas, skin, kidney, brain and resting PBMC. To increase the specificity of the subtraction, the driver was spiked with genes that were determined to be most abundant in the mets3616A cDNA library, such as 15 EF1-alpha, integrin-beta and anticoagulant protein PP4, as well as with cDNAs that were previously found to be differentially expressed in subtracted lung adenocarcinoma cDNA libraries. The resulting subtracted library was referred to as mets3616A-S1.

20 The expression levels of 831 cDNAs from LST-S6 and 521 cDNAs from Mets3616A-S1 in lung tumor tissue and normal tissues was analyzed by microarray technology (Synteni, Palo Alto, CA). Briefly, the cDNAs were PCR amplified and the PCR amplification products were dotted onto slides in an array format, with each product occupying a unique location in the array. mRNA was extracted from the tissue sample to be tested, reverse transcribed, and fluorescent-labeled cDNA probes were generated. The microarrays were probed with the labeled cDNA probes, the slides scanned and 25 fluorescence intensity was measured. This intensity correlates with the hybridization intensity. Thirty-four non-redundant cDNA clones showed 5-fold over-expression in lung tumors, compared with expression in normal tissues tested (lung, skin, lymph node, colon, liver, pancreas, breast, heart, bone marrow, large intestine, kidney, stomach, brain, small

intestine, bladder and salivary gland). The determined cDNA sequences for the 34 isolated clones are provided in SEQ ID NO:77-110.

These sequences were compared to known sequences in the gene bank using the EMBL and GenBank databases. The sequences of SEQ ID NO:77, 86, 90 and 108 were 5 found to show some homology to previously identified expressed sequence tags (ESTs). The sequences of SEQ ID NO:78-85, 87-89, 91-107 and 109-110 were found to show some homology to previously identified genes.

The determined cDNA sequences of 54 clones isolated from lung tumor cDNA libraries that were shown to be differentially over-expressed in non-small cell lung 10 carcinoma by are provided in SEQ ID NO:111-142 and 367-395.

### EXAMPLE 3

#### USE OF PATIENT SERA TO IDENTIFY DNA SEQUENCES ENCODING LUNG TUMOR ANTIGENS

15

This example illustrates the isolation of cDNA sequences encoding lung tumor antigens by expression screening of lung tumor samples with autologous patient sera.

A cDNA expression library was prepared using mRNA from the lung small cell carcinoma cell line NCIH69 in the lambda ZAP Express expression vector (Stratagene) 20 as described above, and screened with a pool of lung small cell carcinoma patient sera. The sera pool was adsorbed with *E. coli* lysate and human PBMC lysate was added to the serum to block antibody to proteins found in normal tissue. Screening was performed as described in Sambrook et al., (*Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989), with the secondary antibody being 25 goat anti-human IgG-A-M (H + L) conjugated with alkaline phosphatase, developed with NBT/BCIP (Gibco BRL). Positive plaques expressing immunoreactive antigens were purified. Phagemid from the plaques was rescued and the nucleotide sequences of the clones was determined.

The determined cDNA sequences of 86 isolated clones are provided in SEQ ID NO:143-228. The sequences of SEQ ID NO:153, 154, 163, 178, 186, 202, 203, 218 and 219 were found to show some homology to previously identified ESTs. The sequences of SEQ ID NO:143-152, 155-162, 164-177, 179-185, 187-201, 204-217 and 220-228 were 5 found to show some homology to previously isolated genes. The sequences of an additional three isolated clones (referred to as SCC2-16, SCC2-28 and SCC2-620 are provided in SEQ ID NO:437-439.

The expression levels of certain of the isolated antigens in lung tumor tissues compared to expression levels in 36 normal tissues was determined using 10 microarray technology and computer analysis. The results of these studies are shown below in Table 3, together with the database analyses for these sequences.

**Table 3**

			Lung Tumor Over-Expression ( $\geq 2$ )			
Clone Name	SEQ ID NO:	Description	LT+F/N	SC+M/N	Squa/N	Aden/N
SCC2-5	146	Unknown KIAA0878	-	3.2	-	2.2
SCC2-9	148	Tubulin K-alpha-1	-	2.2	-	-
SCC2-10	149	NY-REN-64 (pro kinase)	2.0	14.8	3.5	-
SCC2-11	150	TBP-assoc. fact. 2-170	-	5.4	-	-
SCC2-13	152	Centromere Pro F (CENPF)	2.3	5.4	2.8	-
SCC2-14	153	BRUNOL-4	2.7	9.4	3.1	2.3
SCC2-16	437	Non metastatic cells 2	2.3	2.1	2.7	-
SCC2-17	154	Novel (V87915)	2.0	-	2.8	-
SCC2-20	156	Cytoplas Linker Pro 170a2	2.1	3.0+	2.8	-
SCC2-23	157	Hypoxia-induc fact 1 a	2.2	3.0	2.7	-
SCC2-24	158	Actin gamma 1	-	3.8	2.0	-
SCC2-28	438	CHORD-containing pro 1	-	2.1	-	-
SCC2-29	160	Unk. DJ0669I17; ALR-like	2.2	4.0	3.0	2.2
SCC2-31	162	Unknown chrom 1	2.2	3.4	2.8	-
SCC2-36	165	Unknown (T20633)	3.3	2.6	5.5	3.4

SCC2-37	166	Sex-det reg Y Box 21 SOX 2	-	3.0	-	-
SCC2-43	170	CHORD-containing pro 1	-	2.1	-	-
SCC2-50	231	Hypoxia-induc fact 1 a	6.0	3.9	13.7	5.0
SCC2-51	175	Unknown KIAA1051	2.1	3.6	2.0	-
SCC2-54	178	Unknown FLJ20725	-	2.4	2.1	-
SCC2-60	181	Unknown Cosmid R32889	-	3.2	-	-
SCC2-62	439	CHORD-containing pro 1	-	2.1	-	-
SCC2-66	183	Novel, similar to transferase	-	2.2	-	-
SCC2-68	184	Ribosomal pro S7 (RPS7)	-	2.2	-	-

LT+F/N = Lung Tumor plus Fetal tissue over Normal tissues

SC+M/N = Lung Small Cell carcinoma plus Metastatic over Normal tissues

Squa/N = Squamous lung tumor over Normal tissues

5 Aden/N = Adenocarcinoma over Normal tissues

The expression levels of certain other of the isolated antigens in lung tumor tissues compared to expression levels in 36 normal tissues was determined using 10 microarray technology and either computer or visual analysis. The results of these studies are shown below in Table 4, together with the databank analyses for these sequences. These results indicate that these antigens are over-expressed in lung tumor tissue compared to normal tissue.

15

**Table 4**

Clone Name	SEQ ID NO:	Description	Ratio T/N Mean/Med
SCC2-58	180	Sox-2	visual
SCC2-79	190	Mixed-lineage leukemia 4 / AF-6	2.0/2.0 sq
SCC2-91	194	Hepatocellular carcinoma-assoc Ag 58	visual
SCC2-100	200	F-Box protein FBW2	visual

Clone Name	SEQ ID NO:	Description	Ratio T/N Mean/Med
SCC2-102	202	Novel	visual
SCC2-104	204	MAP-kinase act death domain MADD	visual
SCC2-143	214	Unknown HSPC232	2.8/2.1 sq
SCC2-266	226	HMG-2	visual

Ratio T/N = lung tumor tissues over normal tissues

#### EXAMPLE 4

5

#### CLONING OF cDNAs ENCODING LUNG SMALL CELL CARCINOMA ANTIGENS

Lung small cell carcinoma antigens were cloned by screening a small cell cDNA expression library with a mouse anti-SCID mouse serum. This antiserum was developed by growing lung small cell carcinoma cell lines NCIH69 and NCIH128 in SCID 10 mice, removing SCID serum containing shed and secreted tumor antigens and immunizing normal mice with this serum. The library was constructed with mRNA from cell line NCIH128 in the lambda ZAP Express expression vector (Stratagene). The antiserum was adsorbed with *E. coli* lysate and human GAPDH protein and Ku autoantigens, and human PBMC lysate was added to the serum to block antibody to proteins found in normal tissue. 15 Table 5 lists the data bank analyses for the nucleotide sequences. The determined cDNA sequences of the clones are provided in SEQ ID NO:258-317.

**Table 5**

SEQ ID NO.:	Clone ID #	Genbank Homologies
258	54533	Novel
259	54534	Homo sapiens mRNA for LAK-1
260	54536	Homo sapiens CGI-108 protein mRNA
261	54538	Human mRNA for HHR23A protein
262	54540	Homo sapiens chromosome 17, clone hRPC_1030_0_14
263	55084	Homo sapiens homolog of rat elongation factor p18 (p18)
264	55086	Homo sapiens HSPC194 mRNA

SEQ ID NO.:	Clone ID #	Genbank Homologies
265	54555	Homo sapiens accessory proteins BAP31/BAP29 (DXS1357E) mRNA
266	54557	Homo sapiens mesenchymal stem cell protein DSCD75 mRNA
267	54564	Homo sapiens prp28, U5 snRNP 100 kd protein (U5-100K) mRNA
268	55098	Novel
269	55473	Homo sapiens uroporphyrinogen III synthase (congenital erythropoietic porphyria) (UROS)
270	55104	Homo sapiens carbonyl reductase (LOC51181)
271	55105	Homo sapiens membrane component, chromosome 11, surface marker 1 (M11S1)
272	55107	H.sapiens mRNA encoding GPI-anchored protein p137
273	55108	Novel
274	55114	Homo sapiens mRNA; cDNA DKFZp56401716
275	55477	H.sapiens YB-1 gene promoter region
276	55482	Homo sapiens mRNA ; cDNA DKFZp434B0425
277	55483	Human Gu protein mRNA
278	55485	Homo sapiens 45kDa splicing factor mRNA
279	55487	Homo sapiens genomic DNA, chromosome 21q, section 72/105
280	55488	Homo sapiens chromosome 17, clone hCIT529110
281	55087	Novel (partial overlap of Unknown: Homo sapiens partial mRNA, clone c1-10e16)
282	55089	Homo sapiens scaffold attachment factor A (SAF-A) mRNA
283	55092	Homo sapiens density regulated protein drp1 mRNA
284	55093	H.sapiens mRNA encoding GPI-anchored protein p137
285	56926	Homo sapiens high-mobility group (nonhistone chromosomal) protein 17 (HMG17)
286	56930	Novel
287	56944	Homo sapiens KBNA-2 co-activator (100kD) (p100), mRNA
288	56945	Novel
289	55490	Homo sapiens death-associated protein 6 (DAXX) mRNA, and translated products.
290	55495	Homo sapiens mRNA for MEGF6
291	55504	Mus musculus hairy / enhancer of split 6 mRNA
292	55506	Novel / (136bp: Mus musculus mRNA for Rab24 protein)
293	56480	Novel
294	56482	H.sapiens DNA from chromosome 19-cosmids R31158, R31874, & R28125, genomic seq.
295	56484	Novel

SEQ ID NO.:	Clone ID #	Genbank Homologies
296	56487	Human L23 mRNA for putative ribosomal protein
297	56488	Homo sapiens cDNA FLJ10526 fis, clone NT2RP2000931, highly similar to MATRIN 3
298	56490	Homo sapiens Sul1 isolog mRNA
299	56493	Novel
300	56494	Homo sapiens mRNA; cDNA DKFZp564B167 (from clone DKFZp564B167)
301	56495	Homo sapiens 12p13.3 BAC RPCI11-543P15 (Roswell Park Cancer Inst. Human BAC lib.)
302	56499	Human DNA-binding protein B (dbpB) gene, 3' end
303	56517	Homo sapiens esterase D mRNA
304	56952	Homo sapiens 14q32 Jagged2 gene, complete cds; and unknown gene
305	56953	Homo sapiens DNA polymerase zeta catalytic subunit (REV3L) mRNA
306	56959	Novel
307	57139	Homo sapiens ribosomal protein, large, PO (RPLPO) mRNA
308	57078	Homo sapiens alpha-tubulin isoform 1 mRNA
309	57092	Novel
310	57099	Homo sapiens uncharacterized hypothalamus protein HBEX2 mRNA
311	57100	Novel (last 120 bp: Unknown: Canine 21 kDa Signal peptase subunit mRNA)
312	57105	Homo sapiens splicing factor, arginine/serine-rich 7 (35kD) (SFRS7)
313	57111	Human chromosome 14 DNA sequence
314	57117	Human DNA sequence from cosmid V857G56, between markers DXS366 and DXS87 on chromosome X contains ESTs
315	57121	Homo sapiens genomic DNA of 8p21.3-p22 anti-oncogene of hepatocellular colorectal and non-small cell lung cancer, segment 3/11
316	57124	H.sapiens MLN50 mRNA
317	57125	Homo sapiens calreticulin (CALR) , mRNA

EXAMPLE 5

## cDNAs ENCODING LUNG SMALL CELL CARCINOMA ANTIGENS

Lung small cell carcinoma antigens were cloned by screening a small cell

5 cDNA library (NCIH 128) with small cell carcinoma patient sera. The library was constructed with mRNA from cell line NICH 128 in the lambda ZAP Express expression vector (Stratagene). The antiserum was adsorbed with *E. coli* lysate and human GAPDH protein, and human PBMC lysate was added to the serum to block antibody to proteins found in normal tissue. Table 6 lists the homologies identified by database analyses for  
 10 nucleotide sequences shown in SEQ ID NO:318-364. An additional isolated cDNA sequence (referred to as SCC3-90) is provided in SEQ ID NO:440.

Table 6

SEQ ID NO:	Clone ID #	Genbank Homologies
318	54800	Human Ig germline H-chain G-E-A region B
319	54802	Human mRNA for T-cell cyclophilin
320	54803	Unknown BAC clone GS1-11E15
321	54805	Unknown Homo sapiens cDNA FLJ20272 fis
322	54806	Unknown Homo sapiens mRNA for KIAA0713 protein
323	54809	Unknown Homo sapiens mRNA for RIE2 sid2705
324	54810	Homo sapiens glutamyl-prolyl-tRNA synthetase
325	54813	Unknown Human mRNA for KIAA0262 gene
326	54814	Hu.vacuolar proton pump delta polypeptide (VATD) mRNA
327	54816	Unknown Homo sapiens mRNA for KIAA0713 protein
328	54817	Unknown Hu.Chromosome 16 BAC clone CIT987SK-A-101F10
329	54819	Homo sapiens chromokinesin KIF4 (KIF4) mRNA
330	54821	Unknown Homo sapiens cDNA FLJ11101 fis
331	54823	Human mRNA for heat shock protein hsp86
332	54824	hinge=OXPHOS system complex III mitochondrial subunit
333	54825	H.sapiens mRNA for huntingtin interacting protein HIP-I
334	54826	Homo sapiens kinesin light chain mRNA
335	54827	Homo sapiens kinesin light chain mRNA

SEQ ID NO:	Clone ID #	Genbank Homologies
336	54829	Novel
337	54830	Unknown complete sequence
338	54832	Unknown Homo sapiens cDNA FLJ20272 fis
339	55800	Homo sapiens mRNA for E-MAP-115/105
340	55801	Hu. U-snRNP-associated cyclophilin (USA-CyP) mRNA
341	55803	Human chromosome 14 DNA sequence
342	55804	Human thymosin beta-4 mRNA
343	55805	Homo sapiens huntingtin interacting protein 1 (HIP1)
344	55806	Hu. protein kinase, interferon-inducible double stranded RNA
345	55808	Homo sapiens glutathione S-transferase A4 (GSTA4) mRNA
346	55810	Human chromosome 14 DNA sequence
347	55811	Unknown Homo sapiens mRNA for KIAA0713 protein
348	55812	Novel
349	55814	Human poly(ADP-ribose) synthetase mRNA
350	55816	Novel
351	55817	Homo sapiens centromere protein E (CENPE) mRNA
352	55819	Human poly(ADP-ribose) polymerase mRNA
353	55820	Novel
354	55823	Human mRNA for heat shock protein hsp86
355	55824	Novel
356	55826	Homo sapiens SOX18 mRNA, complete cds
357	55828	Novel
358	55829	Novel
359	55831	Unknown BAC sequence from the SPG4 candidate region
360	55832	Homo sapiens heat shock transcription factor 2 (HSF2)
361	55833	Homo sapiens vacuolar H-ATPase subunit D mRNA
362	55834	Homo sapiens clone 628 unknown mRNA
363	55835	Human mRNA for Cu/Zn superoxide dismutase (SOD).
364	55838	Homo sapiens cDNA FLJ20473 fis, clone KAT07092

The expression levels of certain of the isolated antigens in lung tumor tissues compared to expression levels in 36 normal tissues was determined using 5 microarray technology and either computer or visual analysis. The results of these studies are shown below in Table 7, together with the databank analyses for these sequences.

These results indicate that these antigens are over-expressed in lung tumor tissue compared to normal tissue.

**Table 7**

5

Clone Name	SEQ ID NO:	Description	Ratio T/N Mean/Med
SCC3-5	320	Novel	visual
SCC3-7	321	=SCC3-52; Unknown FLJ20272	visual
SCC3-17	325	Ring finger protein 10	2.1/3.0 ad
SCC3-30	330	Unknown FLJ11101	visual
SCC3-52	338	Unknown cDNA FLJ20272	2.7/1.2 sm
SCC3-64	340	U-snRNP-assoc. cyclophilin	visual
SCC3-71	341	TNG-2	visual
SCC3-79	345	GST A4	visual
SCC3-87	349	Poly(ADP-ribose) synthetase	visual
SCC3-90	440	Polyadenylate binding pro (TIA-1)	visual
SCC3-111	359	Unknown BAC	visual
SCC3-112	360	Heat-shock transcription factor 2	visual

Ratio T/N = lung tumor tissues over normal tissues

**EXAMPLE 6**

10

ANALYSIS OF cDNA EXPRESSION USING MICROARRAY TECHNOLOGY

In additional studies, sequences disclosed herein were found to be overexpressed in specific tumor tissues as determined by microarray analysis. Using this approach, cDNA sequences are PCR amplified and their mRNA expression profiles in tumor and normal tissues are examined using cDNA microarray technology essentially as described (Shena *et al.*, 1995). In brief, the clones are arrayed onto glass slides as multiple replicas, with each location corresponding to a unique cDNA clone (as many as 5500 clones can be arrayed on a single slide, or chip). Each chip is hybridized with a pair of cDNA probes that are fluorescence-labeled with Cy3 and Cy5, respectively. Typically, 1 $\mu$ g of polyA<sup>+</sup> RNA is used to generate each cDNA probe. After hybridization, the chips are scanned and the fluorescence intensity recorded for both Cy3 and Cy5 channels. There are

multiple built-in quality control steps. First, the probe quality is monitored using a panel of ubiquitously expressed genes. Secondly, the control plate also can include yeast DNA fragments of which complementary RNA may be spiked into the probe synthesis for measuring the quality of the probe and the sensitivity of the analysis. Currently, the 5 technology offers a sensitivity of 1 in 100,000 copies of mRNA. Finally, the reproducibility of this technology can be ensured by including duplicated control cDNA elements at different locations.

Clones SCC2-5 (SEQ ID NO:229), SCC2-14 (SEQ ID NO:230), SCC2-50 (SEQ ID NO:231) and SCC2-51 (SEQ ID NO:232) were found to be overexpressed by 10 microarray analysis in adenocarcinoma, lung pleural effusion, squamous cell carcinoma, small cell carcinoma, colon tumor, and ovarian tumor, with low levels of expression being detected in all normal tissues tested. The normal tissues included in the microarray were lymph node, salivary gland, lung, bladder, bone marrow, bronchus, esophagus, kidney, heart, liver, lung, skeletal muscle, spleen, stomach, PBMC, skin, thymus, tonsil, trachea, 15 pituitary gland, adrenal gland, brain, pancreas, thyroid gland, adult lung, colon, small intestine, ovary, and peritoneal epithelium. These cDNAs were cloned from a lung small cell carcinoma expression library using small cell carcinoma patient sera as a probe. SCC2-14 has some similarity to an RNA-binding protein, and SCC2-50 is homologous to hypoxia-inducible factor 1 alpha. Amino acid sequences encoded by these cDNAs (SEQ 20 ID Nos:229-232) are shown in SEQ ID NOs:233-236, respectively.

Also by microarray analysis, SCC2-54 (SEQ ID NO:178) was found to be over-expressed in lung small cell and squamous carcinomas relative to normal tissues. An extended cDNA sequence for this clone is provided in SEQ ID NO:396, encoding the polypeptide sequence set forth in SEQ ID NO:397.

25 LSC-49 (SEQ ID NO:29) was found to be overexpressed in lung carcinomas, particularly small cell lung carcinomas. An extended sequence for this clone is provided in SEQ ID NO:412, encoding an amino acid sequence set forth in SEQ ID NO:413. Database searches of LSC-49 revealed sequence homology with a GTPase-activating protein for Rac (mgcRacGAP).

The results of an additional microarray analysis, performed using a criteria of greater than or equal to 2-fold over-expression in tumors and the average expression in normal tissues less than or equal to 0.2 (range from 0.01-10), are summarized in Table 8 below.

5

**Table 8**

Chip #	Clone ID #	Ratio	Mean Signal 1	Mean Signal 2	SEQ ID NO:
5	56908	3.78	0.837	0.221	398, 243
5	56911	2.29	0.453	0.198	399, 245
5	56912	2.57	0.265	0.103	400, 247
5	56913	2.21	0.306	0.138	401, 249
5	56916	2.44	0.449	0.184	402, 251
5	56917	2.29	0.479	0.209	403, 252
5	56921	2.54	0.418	0.165	404, 253
5	56922	5.05	0.613	0.121	405, 255
5	56923	2.74	0.426	0.155	406, 257

The ratio of signal 1 to signal 2 in Table 8 above provides a measure of the level of expression of the identified sequences in tumor versus normal tissues. For 10 example, for SEQ ID NO:398, the tumor-specific signal was 3.78 times that of the signal for the normal tissues tested; for SEQ ID NO:399, the tumor-specific signal was 2.29 times that of the signal for normal tissues, etc.

Results from an additional microarray analysis, performed using visual analysis for identifying cDNAs over-expressed in selected tumor samples, are provided in 15 Table 9 below. Some of these cDNAs were preferentially over-expressed in small cell lung carcinoma (SCLC) samples even though the original cDNAs were identified from subtracted NSCLC tumor samples.

**Table 9**

Chip #	Clone ID #	Ratio	Mean Signal 1	Mean Signal 2	SEQ ID NO:
5	60974	3.84	0.584	0.152	407
5	60976	3.73	0.58	0.155	408
5	60977	3.84	0.492	0.128	409
5	60978	4.63	0.476	0.103	410
5	60980	3.4	0.557	0.164	411

5 In further studies, the expression levels of certain of the isolated antigens in lung tumor tissues previously disclosed in Example 4 were compared to the expression levels in 36 normal tissues using microarray technology and computer analysis. The results of these studies are shown below in Table 10.

**Table 10**

Clone Name	Clone ID #	SEQ ID NO:	Squa/N	Aden/N	SC/N
LSCC2-1	54533	258	3	2	1
LSCC2-2	54534	259	5	3	5
LSCC2-4	54536	260	3	2	2
LSCC2-8	54540	262	0	3	2
LSCC2-18	55084	263	2	2	1
LSCC2-23	54555	265	2	3	3
LSCC2-25	54557	266	2	1	1
LSCC2-32	54564	267	2	3	2
LSCC2-48	55473	269	4	2	1
LSCC2-58	55104	270	3	5	2
LSCC2-61	55107	272	2	5	3
LSCC2-75	55483	277	2	4	2
LSCC2-79	55487	279	3	2	2
LSCC2-93	55089	282	5	4	4
LSCC2-121	55490	289	4	2	2
LSCC2-127	55495	290	2	4	1
LSCC2-137	55504	291	0	3	8
LSCC2-139	55506	292	3	4	1
LSCC2-161	56480	293	3	2	1
LSCC2-164	56482	294	2	4	2
LSCC2-171	56488	297	6	4	5
LSCC2-178	56494	300	3	5	3
LSCC2-191	56517	303	5	2	2

5 Squa/N = Squamous lung tumor over Normal tissues

Aden/N = Adenocarcinoma over Normal tissues

SC/N = Lung Small Cell carcinoma over Normal tissues

**FURTHER CHARACTERIZATION OF THE LUNG TUMOR ANTIGEN L43E**

The predicted protein sequence shown in SEQ ID NO:436 represents a second open reading frame (ORF-2) encoded by the SCC2-51 cDNA nucleotide sequence (also referred to as L43E). The SCC2-51 nucleotide sequence is shown in SEQ ID

NO:175. This protein sequence has 33% identity and 49% similarity to the pol polyprotein of the fish Takifugu rubripes retrotransposon. Motif searches indicate potential protease signatures and protein translocation analysis indicates that the protein could be cytoplasmic or membrane-associated due to a potential transmembrane region. Using realtime PCR, 5 SCC2-51 was found to be over-expressed in primary small cell carcinoma and in atypical carcinoid metastatic tumors, but weakly expressed in other lung carcinomas and normal tissues except for pituitary gland and adrenal gland. The cDNA sequence and ORF-1 have homology to Takifugu rubripes gag polyprotein (28% identity and 45% similarity).

10

#### EXAMPLE 8

##### ISOLATION OF cDNA SEQUENCES FOR ADDITIONAL LUNG TUMOR ANTIGENS

Additional cDNA clones were obtained from analysis II of LST-S6 and Mets3616-S1 libraries of Lung Chip V. These cDNAs were differentially expressed in lung squamous and/or adenocarcinoma tumors (greater than or equal to 2 fold), and the average 15 expression values for these clones in normal tissues were below 0.1 (the range of value was between 0.001 and 10). A total of 29 non-redundant cDNA sequences were isolated and are disclosed in SEQ ID NO:367-395. A summary of these clones with respect to the Genbank searches is shown in Table 11.

**Table 11**

<u>SEQ ID NO:</u>	<u>Clone ID #</u>	<u>Chip #</u>	<u>GenBank</u>
367	49949	5	Novel
368	49952	5	Collagen type IV alpha-5
370	49960	5	h. mRNA for Pirin, isolate 17
371	49961	5	vector/Novel
372 and 373	49962	5	HBP, heme binding protein
374	49965	5	h. testitin
375	49966	5	KIAA 1077
376	49977	5	Cyclin B homologue
377	49975	5	Cat Eye 22q11.2
378	49982	5	Novel
379	49986	5	Novel
380	49988	5	KIAA0292, similar to AR1 protein
381	49993	5	transferrin receptor
382	49995	5	Cathepsin B
383	49996	5	RP3, similar to mouse tctex-1
384	49999	5	Novel, Cosmid g1572c198
385	50006	5	sheep and mouse sox2 gene (HMG box, germ cell)
386	50007	5	Nrf3 for NF-E2 related factor 3
387	50009	5	vector/Novel, chrom. 10 seq
388	50014	5	clone RP5-1025A1 on 20p11.21
389	50016	5	Failed/h. MEGF9
390	50017	5	NH0160k17
391	50019	5	None
392	50022	5	h mitotic kinesin-like protein-1
393	50023	5	KIAA1077
394	50024	5	None
395	50033	5	None

### EXAMPLE 9

#### REAL-TIME PCR ANALYSIS OF L578S

As previously shown in Example 2, clone 48137 (SEQ ID NO:89), which is 5 also referred to as L578S, and is predicted to have an extended cDNA sequence of SEQ ID NO:365, was shown to be 5-fold over-expressed in lung tumors as compared to the normal tissue by microarray analysis. Real-time PCR analysis confirmed that L578S is over-expressed in both lung squamous and adenocarcinoma tumors. Database analysis identified two human proteins showing some degree of homology to L578S, one corresponding to a 10 putative type Ib membrane-bound protein. Protein alignment between this protein and SEQ ID NO:365 indicated that L578S full-length protein may also be a type Ib membrane-protein. This indicates that L578S is an attractive target for the development of antibody-based therapeutics.

15

### EXAMPLE 10

#### SYNTHESIS OF POLYPEPTIDES

Polypeptides are synthesized on a Perkin Elmer/Applied Biosystems 20 Division 430A peptide synthesizer using Fmoc chemistry with HPTU (O-Benzotriazole-N,N,N',N'-tetramethyluronium hexafluorophosphate) activation. A Gly-Cys-Gly sequence is attached to the amino terminus of the peptide to provide a method of conjugation, binding to an immobilized surface, or labeling of the peptide. Cleavage of the peptides from the solid support is carried out using the following cleavage mixture: trifluoroacetic acid:ethanedithiol:thioanisole:water:phenol (40:1:2:2:3). After cleaving for 2 hours, the peptides are precipitated in cold methyl-t-butyl-ether. The peptide pellets are then dissolved in water containing 0.1% trifluoroacetic acid (TFA) and lyophilized prior to purification by C18 reverse phase HPLC. A gradient of 0%-60% acetonitrile (containing 25 0.1% TFA) in water (containing 0.1% TFA) is used to elute the peptides. Following

lyophilization of the pure fractions, the peptides are characterized using electrospray or other types of mass spectrometry and by amino acid analysis.

From the foregoing it will be appreciated that, although specific 5 embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention. Accordingly, the invention is not limited except as by the appended claims.

127